

CIRCUITS AND SYSTEMS FOR WIRELESS CONCURRENT COMMUNICATION

Thesis by

Yu-Jiu Wang

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2009

(Defended February 26, 2009)

© 2009

Yu-Jiu Wang

All Rights Reserved

To My Grandparents and Parents

Acknowledgements

I would like to express my sincerest gratitude to my research advisor, Professor Ali Hajimiri, for his excellent guidance, encouragement, and patience during the years of my Ph.D. Ali is not only my advisor, but also my role model. From him, I have learned how to improve myself consistently, how to treat my promises seriously, how to dedicate myself to my profession, and how to think and live independently. I am in particular grateful to his tolerance of the many troubles I made for him in the past four years and five months.

I would like to thank Professor Sander Weinreb for his technical support and assistance over the course of my Ph.D. In particular, I am grateful for his technical inputs during the concurrent phased array project and the modified FET noise models project.

I would also like to thank Professor Ali Hajimiri, Professor Sander Weinreb, Professor Dave Rutledge, Professor Azita Emami, and Professor P.P.Vaidyanathan for serving on my candidacy and defense committees.

I would like to thank John DeFalco, Mike Sarcione, and Richard Healy of Raytheon Company for providing us a loaded research fund for the concurrent phased array project. I would also like to thank Rahul Magoon and David Kang in Axiom Microdevices for providing us a generous shuttle run. I am grateful to Shohei Kosai from Toshiba for providing us a shuttle run. We spent the most unforgettable New Year's Eve of my life in the lab, and I am impressed by his work ethics.

I would like to thank all members of Caltech High Speed Integrated Circuits, Caltech Millimeter-Wave IC, and Caltech Mixed-Mode Integrated Circuits and Systems groups. I am particularly thankful to Dr. Aydin Babakhani for making the early part of my Ph.D. “extraordinary”, as well as providing many interesting ideas regarding both research and lifestyles. I would also like to thank Hua Wang, for discussing many technical questions with me, and helping me with many experiments and designs. He is also a good friend to talk about different general issues, and to have “ordinary” fun. I would also like to thank Dr. Xiang Guan, Dr. Arun Natarajan, Dr. Behnam Analui, Dr. Abbas Komijani, Professor Ehsan Afshari, Sam Mandegaran, Professor Jim Buckwalter, Professor Arjang Hassibi and Dr. Chris White for their mentoring during the early part of my Ph.D. I wish to thank my colleagues Professor Sangguen Jeon, Florian Bohn, Edward Keehr, Juhwan Yoo, Jay Chen, Jennifer Arroyo, Kaushik Sengupta, Steven Bowers, Kaushik Dasgupta, Alvaro Gonzales, and Tomoyuki Arai from Fujitsu for their support. I also wish to thank Yu-Lung Tang, Matthew Loh, and Joe Bardin in the fourth floor for their help.

I would like to thank Michelle Chen, Dale Yee, Naveed Near-Ansairi, John Lilley, Hamdi Mani, Niklas Wadefalk, Ann Shen, Linda Dozsa, Tanya Owen, Carol Sosnowski, Gary Waters, Lynn Hein, Janet Couch, and Kent Porter for their valuable assistance during my time at Caltech.

I am very fortunate to have met many friends in/around Caltech, and they made my Ph.D. a very pleasant journey. They are Josephine Wang (and her pet dog Bowie Wang), Howie and Derry Ge, Hua’s fiancé Ying Wang, Yan Chen, Teddy Yu, Yu-Teng Chang, Chun-Yang Chen, Dr. Borching Su, Ching-Chih Weng, Dr. Po-Jui Chen, Dr. Wei-Hsin Gu,

Professor Chin-Lin Guo, Ray Huang, Professor Hsuan-Tien Lin, Chun-Hui Lin, Mingshir Lin, Sebastien Lasfargues, Shafigh Shirinfar, John Howard, and Richard Ohanian.

I would also like to thank Professor Huei Wang at National Taiwan University who brought me to this exciting field ten years ago, so I would have the chance to write this thesis. I would also like to thank Professor Ming-Juey Lin in National Taiwan Normal University, who chose me to the International Physics Olympiad national team, and showed me how large the world is during my high school years.

Finally, I thank my grandparents and my parents for providing me a family that appreciates knowledge, a life without financial worries, and giving out their love to me without reserve. I also thank my sisters for their constant encouragement during my studies.

Abstract

Concurrency is a special kind of analog circuit parallelism that uses a single circuit with necessary bandwidth to process multiple signals at the same time. Concurrent radios offer a higher data rate and improved system diversity. Our comprehensive treatment comprises proposals for potential transceiver architectures, invention of circuit blocks, and provisions of innovative analysis methods.

The analysis of concurrent circuits are often complex. To simplify noise analysis, a R^{N^2} -vector space is first proposed to re-formulate the N-port network noise modeling problem. Any internal physical source inside the noisy network contributes a small vector in the defined R^{N^2} -vector space, and the aggregate statistical behavior of this noisy network can be viewed as the vector sum of these vectors. Applying this concept to FET noise modeling leads to several modified FET noise models, in which three uncorrelated noise sources are sufficient to describe the statistical behavior of an intrinsic FET. The use of these new FET models can simplify the analysis, simulation, and optimization of low noise systems without sacrificing accuracy.

Broadband low-noise amplifier is a critical block in concurrent receiver systems. We propose a novel low-noise weighted distributed amplifier (WDA) topology, which uses the internal finite-impulse-response filtering inside a conventional distributed amplifier to partially suppress internal thermal noise. A distinct advantage of this topology is its tolerance to input parasitic capacitance which can be used to provide good electro-static discharge (ESD) protection without sacrificing its noise performance and power

consumption. A compact 3.1—10.6 GHz WDA IC is built on a 130 nm CMOS process.

Experimental results show 2.3—4.5 dB NF at 23 mW power consumption.

Using concurrency in wireless link can boost communication data rate. As a proof-of concept, we propose dynamically scalable concurrent communication by dividing the 7.5 GHz bandwidth of the unlicensed 3.1—10.6 GHz spectrum into seven concurrent channels. A CMOS octa-core RF receiver is implemented to validate the idea. Based on the receiver measurement results, a wireless link can be built to achieve a 16 Gbps channel limit at five meter TX-RX distance at 400 mW power consumption.

Tunable concurrency can improve the receiver diversity. A prototype 6—18 GHz concurrent tunable dual-band phased array receiver element IC is proposed and built on a 130 nm CMOS process. Experimental results demonstrate successful dual-band RF reception within a low band (6—10.4 GHz) and high band (10.4—18 GHz) with 300 MHz baseband bandwidth. A final four-element phased array receiver built from the prototyped ICs shows an array pattern with worst-case 21 dB peak-to-null ratio across all frequencies.

Concurrency can also be used to achieve multi-beam reception by providing multiple phase-shifts for each RF signals and combining them separately at baseband outputs. A 10.4—18 GHz concurrent dual-beam phased array receiver is proposed based on this concept, and is implemented on a 130 nm CMOS process. A final four-element phased array system shows successful concurrent dual-beam reception at the same RF frequency.

Table of Contents

Acknowledgements	iv
Abstract	vii
Table of Contents	ix
Chapter 1: Introduction	1
1.1. Organization	2
Chapter 2: Noisy Network Modeling Using Only Uncorrelated Sources..	5
2.1. Circuit Theory of Linear Noisy Network	7
2.2. Defining Vector Space for Linear Noisy Network	9
2.3. Example: A Two-Port Noisy Network	15
2.4. A Modified FET Noise Model	20
2.5. FET Noise Model Comparisons	23
2.6. Summary	27
Appendix 2.1: Van der Ziel's FET Noise Model	28
Appendix 2.2: Pospieszalski's Noise Model	35
Appendix 2.3: BSIM4 Noise Model	36
Chapter 3: A Compact Low Noise Weighted Distributed Amplifier	37
3.1. Input Matching versus Bandwidth	38
3.2. Issues of Power-Constraint LNA Optimization in CMOS	40
3.3. Low-Noise Distributed Amplifier	42
3.4. Noise Process in the Weighted Distributed Amplifier (WDA)	45

	x
3.4.1. Noise from Common-Source Transistors	46
3.4.2. Noise from Cascode Transistors	49
3.4.3. Noise from Termination Resistors	51
3.4.4. Noise from Passive Network Loss	52
3.4.5. Voltage Peaking Effect in LC-Ladder.....	54
3.4.6. Frequency-Dependent Group Delay	56
3.4.7. Frequency-Dependent Impedance Change	57
3.4.8. Noise Figure of WDA.....	58
3.5. Power-Constraint Noise Optimization of WDA	59
3.6. Magnetic Couplings in LC-Ladder.....	61
3.7. WDA Schematics and Layout	64
3.8. WDA Measurement Results	68
3.9. Summary	72
Appendix 3.1: Integrated RF VLSI Design Flow	73
Chapter 4: Concurrent Octa-Core RF Receiver Architecture	77
4.1 Introduction	78
4.1.1 Wireless Multi-Gbps Communication	78
4.1.2 Comparisons between the 3.1—10.6 GHz and the 60 GHz Band.....	80
4.1.3 Previous Works using the 3.1—10.6 GHz Band.....	80
4.2 A 3.1—10.6 GHz Octa-Core Receiver	82
4.2.1 System Architecture	82
4.2.2 RF Common Part: LNA, RF Buffers, and RF Distribution Network	86

4.2.3 Downconversion Core: PLL, Mixers, and BB Buffers	88
4.2.4 Experimental Results	91
4.3 Summary	99

Chapter 5: Scalable Concurrent Dual-Band Phased Array Receiver ... 100

5.1. Introduction of Phased Array Receiver	101
5.1.1. Limitations of Previous Works on Phased Array	103
5.1.2. Previous Works on Concurrent Dual-Band Receivers	105
5.1.3. Proposed Large-Scale Phased Array System Architecture	105
5.2. Tunability of Concurrent Dual-Band Amplifiers	108
5.3. Tunable Concurrent Amplifier (TCA)	112
5.3.1. Common-Gate Common-Gate (CG-CG) Topology	113
5.3.2. Common-Gate Common-Source (CG-CS) Topology	114
5.3.3. Resistor Termination Topology	116
5.3.4. Active Termination Topology	117
5.4 A 6—18 GHz Concurrent Tunable Dual-Band Phased Array Receiver	121
5.4.1 Block Diagrams	121
5.4.2 TCA	124
5.4.3 RF and IF Mixers	125
5.4.4 Baseband Buffers	127
5.4.5 Whole Receiver Chip	128
5.5 Experimental Results	129
5.5.1 Receiver Test Circuits	129

5.5.2	Four-Element Phased Array Pattern.....	134
5.6	Summary	138
Chapter 6: Concurrent Co-Channel Dual-Beam Phased Array Receiver		
.....		139
6.1.	Dual-Beam Phased Array System Architecture.....	140
6.2.	A 10.4—18 GHz Concurrent Quad-Beam Phased Array Receiver	142
6.2.1.	Receiver Element Block Diagrams	142
6.2.2.	LNA.....	144
6.2.3.	IF Signal Distribution Networks	146
6.2.4.	RF/IF Mixers and Baseband Buffers.....	146
6.2.5.	Receiver Element Implementation	146
6.3.	Experimental Results	148
6.3.1.	Receiver Element Measurement Results.....	148
6.3.2.	Four-Element Phased Array Measurement Results.....	150
6.4.	Summary	156
Chapter 7: Conclusion.....		158
7.1	Summary	158
Bibliography		161

List of Figures

Figure 2.1: Different noisy networks might be able to reduce to the same compact network form.....	9
Figure 2.2: Noise contributions from the internal physical noise sources to the external world can be interpreted as the sum of several noise vectors in the defined vector space.	13
Figure 2.3: Two correlation admittances are used to decorrelate the two noisy sources.	16
Figure 2.4: A controlled source is used to implement a noisy two-port network.....	16
Figure 2.5: (a) A physical two-port linear noisy network, (b) Compact form of a two-port noisy network, (c) A two-port network with four uncorrelated noise sources, and (d) The conceptual plots of the noise vectors for (a) and (c) in an R^4 -space.....	18
Figure 2.6: Three frequency-independent and uncorrelated sources are used to implement Van der Ziel's FET noise model.....	21
Figure 2.7: Another modified FET noise model is also equivalent to Van der Ziel's FET noise model.....	23
Figure 2.8: Pospiezalski's model, BSIM4 model, and modified FET noise model are fitted to the long-channel Van der Ziel's model.	25
Figure 2.9: FET noise modeling is viewed as vector summation in the normalized vector space.	27
Figure 2.10: Linear voltage perturbation distribution along the channel of a FET due to a noise perturbation at x_0	28
Figure 2.11: Direction of the current sources	31

Figure 3.1: Signal strength in wireless communication	37
Figure 3.2: Trade-offs for input matching and input parasitic capacitance	40
Figure 3.3: Transconductance current efficiency and f_T versus bias voltage of a CMOS transistor and an intrinsic BJT	42
Figure 3.4: The concept of distributed amplification	43
Figure 3.5: The concept of weighted distributed amplifier	45
Figure 3.6: Noise from the drain noise of the i -th common-source transistor	47
Figure 3.7: Noise from the gate noise of the i -th common-source transistor	48
Figure 3.8: Noise from the source noise of the i -th common-source transistor	48
Figure 3.9: Noise from the drain noise of the i -th cascode transistor	49
Figure 3.10: Noise from the gate noise of the i -th cascode transistor	50
Figure 3.11: Noise from the source noise of the i -th cascode transistor	51
Figure 3.12: Equivalent circuit of the LC-ladder with a lossy inductor between the $(i-1)$ -th and i -th stage	52
Figure 3.13: Driving the LC-ladders with a broadband power source, and its equivalent circuit	55
Figure 3.14: Driving the output LC-ladder from an internal load, and its equivalent circuit	56
Figure 3.15: Power-constraint noise optimization contour comparisons between a DA and a WDA	59
Figure 3.16: Adjacent couplings in a LC-ladder	62
Figure 3.17: Inductor layouts in two different LC-ladders: (a) Non-alternating inductor layout, and (b) Alternating inductor layout	63
Figure 3.18: Schematics of the WDA	65

Figure 3.19: Schematics of the intermediate amplifiers and their device sizing	65
Figure 3.20: Schematics of the variable termination resistors	66
Figure 3.21: Die micrograph of the WDA	67
Figure 3.22: S-parameters measurement results	68
Figure 3.23: Noise figure (NF), input-referred third-order intercept point (IIP3), and input-referred 1 dB gain compression point (P1dB) measurement results at 26 mW ..	69
Figure 3.24: Measured noise figure (NF) and simulated NF of different transistor γ at 17 mW power consumption	70
Figure 3.25: Worst-case measured performance versus power consumption	71
Figure 3.26: Conventional RF IC design flow	73
Figure 3.27: An integrated RF VLSI design flow	75
Figure 4.1: Previous works on the 3.1—10.6 GHz band: (a) Impulse-band and (b) Frequency-hopping based	81
Figure 4.2: Dynamically scalable concurrent communication	82
Figure 4.3: System architecture of proposed octa-core RF receiver	84
Figure 4.4: Schematics of the RF common part	88
Figure 4.5: Block diagram of the down-conversion core	90
Figure 4.6: Chip micrograph of the octa-core receiver IC	91
Figure 4.7: Schematics of mixer and RF/LO-I buffers inside each RX core, and the results of system healing at a typical VGA gain setting	92
Figure 4.8: Measured receiver maximum conversion gain and S_{11}	94
Figure 4.9: Measured receiver system noise figure and IIP3	94

Figure 4.10: Measured cross-band rejection	95
Figure 4.11: Measured LO spurs and core-to-core LO leakage	96
Figure 4.12: Channel capacity of a wireless link built with the octa-core receiver, with a transmitter transmitting at FCC's spectrum mask and isotopic antennas for both RX and TX	97
Figure 5.1: Basic phased array receiver configuration	101
Figure 5.2: A conventional way of building a large-scale phased-array receiver system (in the active array configuration) that supports concurrent multiple beams.....	104
Figure 5.3: A proposed 6–18 GHz phased array receiver system that receives four beams at two frequencies concurrently and is easily scalable toward a very large-scale array	106
Figure 5.4: Schematics of a concurrent dual-band amplifier	109
Figure 5.5: Achievable frequency region of tunable dual-band operation of the amplifiers in Figure 5.4 with limited capacitor tuning range, and all frequencies between 6— 18 GHz covered by either band	111
Figure 5.6: Schematics of a common-gate common-gate TCA	112
Figure 5.7: Schematics of a common-gate common-source TCA	114
Figure 5.8: Schematics of a resistor-terminated TCA	116
Figure 5.9: Schematics of an active-termination TCA	117
Figure 5.10: Architecture of the tunable concurrent dual-band quad-beam phased array receiver in CMOS.....	122
Figure 5.11: Frequency scheme	123

Figure 5.12: Schematic of the TCA with a single input and a dual output	123
Figure 5.13: Schematic of the RF mixer and IF buffer for LB	125
Figure 5.14: Schematic of the RF mixer for HB.....	126
Figure 5.15: Schematic of baseband VGA	127
Figure 5.16: Chip micrograph	128
Figure 5.17: Block diagram of the receiver measurement setup.....	129
Figure 5.18: Measured input-matching performance	130
Figure 5.19: Measured conversion gain.....	131
Figure 5.20: Measured nonlinearity performance: Input-referred IP3 and 1 dB compression.....	132
Figure 5.21: Measured noise figure of the CMOS receiver (solid line with markers) and the complete system including the active antenna module (dashed line).....	132
Figure 5.22: Measured isolation performance: Cross-band and cross-polarization rejection ratios	133
Figure 5.23: Photo of the four-element array.....	134
Figure 5.24: Electrical array test setup.....	135
Figure 5.25: Measured array patterns of the four-element array. Theoretical patterns are superimposed.....	136
Figure 6.1: Architecture of the 10.4 – 18 GHz co-channel dual-beam phased array receiver system	140
Figure 6.2: Architecture of the tunable co-channel dual-beam phased array receiver element in CMOS (10.4—18 GHz).....	143
Figure 6.3: Schematic of the LNA	145

Figure 6.4: Schematic of the tunable amplifier.....	145
Figure 6.5: Chip micrograph of the 10.4 —18 GHz dual-beam receiver element	147
Figure 6.6: Measured conversion gain and input-matching performance of the 10.4—18 GHz dual-beam receiver element	148
Figure 6.7: Measured input-referred 1 dB gain compression and the IP3 of the 10.4—18 GHz dual-beam receiver element.	149
Figure 6.8: Measured noise figure of the 10.4—18 GHz dual-beam receiver element.....	150
Figure 6.9: Concurrent co-channel dual-beam feed with different DOAs.....	151
Figure 6.10: Measured concurrent dual-beam array patterns at 17.85 GHz of the 10.4—18 GHz co-channel dual-beam phased array. The beam-pointing angle for beam 1 (dashed line) is fixed at 0° . The beam-pointing angle for beam 2 (solid line) is steered at (a) -60° , (b) -30° , (c) 30° , (d) 60° . The antenna spacing is assumed as a half wavelength of the incoming signal.	152
Figure 6.11: Measured cross-beam rejection performance ($f_{RF} = 17.85$ GHz). The incident angle of the desired signal is fixed at 0°	152
Figure 6.12: Desensitization of the array system ($f_{RF} = 17.85$ GHz)	154
Figure 6.13: Measured EVM of the concurrent dual-beam signals, each independently modulated with 4.5 Msps QPSK at 17.85 GHz. The incident angle of beam 1 is fixed at 0°	155

List of Tables

Table 2.1: Comparisons between FET noise models	24
Table 4.1: Measured performance summary of the octa-core receiver.....	98
Table 5.1: Measured performance summary of the scalable concurrent dual-band phased array receiver	137
Table 6.1: Measured performance summary of the concurrent co-channel dual-beam phased array receiver.....	157

Chapter 1: Introduction

In the history of integrated circuits, there have been so many times that people doubt its future: “Are we approaching the physical limit of lithography?”; “Will gate leakage current stop us from scaling?”; “Will parasitics from metal interconnection dramatically degrade the performance of an advanced process?”; “Will IC become too complex for designers to handle in limited time?”; “Will electronics stop improving/evolving?”; etc. Technological innovations like optical-proximity correction, phase-shift mask, strain silicon, high-K gate oxide, metal gate, low-K dielectric, VLSI synthesis, fast-SPICE algorithm [60], etc., have been invented at a convenient time to solve these issues. At the moment this thesis is written, it is fortunate to see this industry continue to roll at its projected speed without any sign of slowing down. It is the creativity and hardwork of scientists and engineers that expand the frontier of technologies.

The continuing improvement of semiconductor technology also enables the advancement of wireless communication electronics. The ability to achieve higher transistor switching speed, higher system-integration and complexity levels offers both design opportunities and challenges to communication engineers to explore and develop innovative IC and products. In consumer markets, we witnessed the burgeoning of pagers give way to the overwhelming rule of cell phones in the last two decades. In military and academia, bulky radar systems made from discrete modules have been integrated into

single-chip silicon-based solutions that do the same tasks. It is exciting to expect more wireless concepts, products, and applications in the near future.

Two major challenges in wireless broadband communication are how to increase system diversity and how to improve broadband radio spectrum efficiency. In this thesis, we will present a unique view on solving these challenges by using concurrency in analog/RF frontend circuitry. Concurrency is a special type of analog circuit parallelism that uses a single circuit with necessary bandwidth to process multiple signals at a same time. Our treatment comprises of the definition of such novel radios, formulation of their particular characteristics, proposals for potential transceiver architectures, invention of circuit blocks, and provisions of innovative analysis methods. Throughout the discussions, our theoretical findings are verified with experimental implementation of the developed concepts.

The contributions of our study include the development of original concepts and new theoretical findings together with practical implications in the area of integrated broadband concurrent multi-band radio systems.

1.1. Organization

This thesis is dedicated to the study of circuits and systems for wireless concurrent communication in the context of RF/Analog circuitry. Chapter 2 and Chapter 3 will emphasize circuit-level problems, research and solutions. The analysis of concurrent circuits are often complex, and simplification will be a necessary step to analyze them. In Chapter 2, we will review the general N-port noise modeling problem, which is a common

problem for low-noise system design. A vector space for a general noisy N-port is proposed to visualize the noise modeling process as series of vector summation. A general noisy two-port is used as an example to further explain the idea. Applying the noisy two-port to the modeling of intrinsic FET leads to several possible modified FET noise models, in which three uncorrelated noise sources are sufficient to describe the statistical behavior of an intrinsic FET. A comparison between the proposed modified FET noise model, Van der Ziel's noise model, Pospieszalski's noise model, and the BSIM4 model is also presented.

Low-noise amplifier (LNA) is a critical building block in wireless concurrent communication. In Chapter 3, we propose the low-noise weighted distributed amplifier (WDA) topology. A distinct advantage of this topology is its tolerance to input parasitic capacitance which can be utilized to provide electro-static discharge (ESD) protection without sacrificing its noise performance and power consumption. The proposed modified FET noise model is applied to simplify noise analysis, simulation, and optimization of the design of a 3.1—10.6 GHz WDA, and a compact test IC is built on a 130 nm CMOS process. Experimental results will be presented which verify the design.

Chapters 4, 5 and 6 are focused on system-level research. In Chapter 4, we will present the use of concurrency to boost communication data rate. As a proof-of-concept, we propose dynamically scalable concurrent communication by dividing the 7.5 GHz bandwidth of the unlicensed 3.1—10.6 GHz spectrum into several concurrent channels. A CMOS octa-core RF receiver is implemented to verify the concept. Measurement results of this receiver will be provided, which indicate that a wireless link can be built based on

this architecture to achieve a 16 Gbps channel limit at five meter TX-RX distance at 400 mW power consumption.

Chapter 5 and Chapter 6 apply concurrency in phased array systems to increase its diversity. Chapter 5 introduces the scalable concurrent tunable dual-band phased array receiver. Design challenges against achieving concurrent tunable dual-band RF signal reception will be studied first, and their alternative solutions will be discussed. A prototype 6–18 GHz receiver element IC is implemented on a 130 nm CMOS process. Experimental results of a single receiver element as well as a final four-element phased array receiver will be demonstrated.

A phased array receiver can achieve spatial filtering at the system output; however, it should be noted that information from different incoming angles are intact before the combining of phase-compensated receiver array outputs. Chapter 6 introduces a concurrent multi-beam phased array receiver which utilizes this property to achieve concurrent multi-beam reception. This topology allows us to share the antenna, RF frontend, and LO circuitry. A prototype receiver IC has been implemented and measured to verify the concept. A final four-element phase array receiver is built based on the receiver IC which proves the possibility of concurrent multi-beam reception.

Last but not least, a summary of the thesis highlights will be given in Chapter 7 to conclude this thesis.

Chapter 2: Noisy Network Modeling Using Only Uncorrelated Sources

Thermal fluctuations of electric charges inside all conductors generate a measurable physical electrical potential between any two ends of the conductors. This random potential was first observed by Johnson in experiments [1], and later Nyquist postulated a black-body radiation thought experiment to relate its noise voltage power to the resistivity of the conductor. Based on Nyquist's derivations, the average noise power of the conductors is $\overline{V_N^2} \approx 4kT\Delta FR$ for $f \ll \frac{kT}{h}$ [2]. Here, k is the Boltzmann constant, T is the temperature of the resistor, ΔF is the measurement bandwidth, R is the resistivity, f is the frequency of noise in concerns, and h is Planck's constant. If an electrical experiment is carried at room temperature ($T = 300K$), $\frac{kT}{h} = 6.24 THz$ suggests $\overline{V_N^2} \approx 4kT\Delta FR$ holds for microwave and millimeter wave ranges.

Since all electronics are built on circuit networks, which are composed of different elementary functional blocks (like resistors, inductors, transistors, etc.) with conducting wires, this electrical noise is an inevitable part of any physical electronics system. In other words, electrical signals processed by any electronics systems will be accompanied with this background thermal noise. For the electronics to work properly, a minimum signal-to-ratio requirement has to be met. This suggests that thermal noise places a lower bound on the dynamic range of any electronics system.

For all practical purposes, either time-invariant or time-variant, noise can be viewed as a small signal deviation from the case when noise is absent. Thus, linearization around the operating point is usually utilized to simplify the noise analysis. Circuit theory of linear noisy networks has been thoroughly studied by Haus and Adler for more than fifty years [3]. If so, why would it be worth it to us to dedicate one chapter in this thesis to discuss it?

Classical noisy network theory is compact in its mathematical form; however, this compactness makes arbitrary noisy networks difficult to implement in electronic design automation (EDA) tools. In Section 2.1, we will briefly review classical circuit theory of linear noisy network. In Section 2.2, we will look at the classical problem from a new perspective by defining vector space for arbitrary noisy networks. Once we do so, it becomes clear that there several possible noisy networks equivalent to the same noisy network, and we can choose the one that is easiest for EDA implementation to use for noise modeling. In Section 2.3, we will show a general two-port noisy network example. We then apply this two-port example for modeling a noisy FET in Section 2.4. This gives rise to several equivalent modified noise models for FET. In Section 2.5, we will compare our proposed models to three other commonly used FET noise models, namely: the Van der Ziel's model, Pospieszalski's model, and the BSIM4 holistic noise model. We summarize main points of this chapter in Section 2.6.

2.1. Circuit Theory of Linear Noisy Network

For any arbitrary N-port linear network with internal independent sources, output signals consist of the parts that are linearly proportional to the input signals and the other parts contributed by the internal independent sources. Without the loss of generality, we can express this input-output relationship using the admittance matrix in frequency domain:

$$\mathbf{I} = \mathbf{Y}\mathbf{V} + \mathbf{I}_s. \quad (2.1.1)$$

\mathbf{V} is the Laplace-transformed input voltages vector: $\mathbf{V} = [v_1(s) \ v_2(s) \ \dots \ v_N(s)]^T$. \mathbf{I} is the Laplace-transformed output currents vector: $\mathbf{I} = [i_1(s) \ i_2(s) \ \dots \ i_N(s)]^T$. \mathbf{I}_s is the Laplace-transformed output currents vector due to the independent sources: $\mathbf{I}_s = [i_{s1}(s) \ i_{s2}(s) \ \dots \ i_{sN}(s)]^T$. Superscript operator $[\]^T$ denotes the transpose of a matrix [4]. Laplace-transform of a time domain signal is defined as [5]:

$$e_m(s) = \int_{0^-}^{+\infty} e_m(t) \cdot e^{-st} \cdot dt. \quad (2.1.2)$$

\mathbf{Y} is the Laplace-transformed admittance matrix:

$$\mathbf{Y} = \begin{bmatrix} y_{11}(s) & y_{12}(s) & \dots & y_{1N}(s) \\ y_{21}(s) & y_{22}(s) & \dots & y_{2N}(s) \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1}(s) & y_{N2}(s) & \dots & y_{NN}(s) \end{bmatrix}, \quad (2.1.3)$$

with its matrix element $y_{mn}(s) \equiv \frac{\partial i_m(s)}{\partial v_n(s)}$.

If we apply the inverse Laplace-transform to Equation (2.1.1), we will get the time-domain representation of the linear network:

$$\mathbf{V}(t) = \mathbf{Y}(t) * \mathbf{I}(t) + \mathbf{I}_s(t). \quad (2.1.4)$$

Here, $\mathbf{V}(t) = [v_1(t) \ v_2(t) \ \dots \ v_N(t)]^T$, $\mathbf{I}(t) = [i_1(t) \ i_2(t) \ \dots \ i_N(t)]^T$, and $\mathbf{I}_s(t) = [i_{s1}(t) \ i_{s2}(t) \ \dots \ i_{sN}(t)]^T$. And the time-domain admittance matrix will become:

$$\mathbf{Y}(t) = \begin{bmatrix} y_{11}(t) & y_{12}(t) & \dots & y_{1N}(t) \\ y_{21}(t) & y_{22}(t) & \dots & y_{2N}(t) \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1}(t) & y_{N2}(t) & \dots & y_{NN}(t) \end{bmatrix}. \quad (2.1.5)$$

The matrix elements of $\mathbf{V}(t)$, $\mathbf{I}(t)$, $\mathbf{I}_s(t)$, and $\mathbf{Y}(t)$ are the time-domain representations of the original matrix elements. The $*$ symbol in Equation (2.1.4) is the matrix convolution operator defined as:

$$\mathbf{A}_{M \times K} * \mathbf{B}_{K \times N} \equiv \left(\sum_{k=1}^K \int_{0^-}^{t^+} a_{mk}(\tau) \cdot b_{kn}(t - \tau) \cdot d\tau \right). \quad (2.1.6)$$

In a linear noisy network modeling problem, these independent sources' contribution to the outputs are random processes. In general, arbitrary random processes are complex to describe. Fortunately, in the case of electronic thermal noise, wide-sense stationary (WSS) property is held. The statistical behavior of WSS random processes can be fully described by their autocorrelation and cross-correlation functions, which are defined as [6]:

$$\mathbf{C}(\tau) = E\{I_s(t) \overline{I_s^T(t + \tau)}\} = (c_{mn}) \quad (2.1.7)$$

$$c_{mn}(t) = E\{i_m(t) \overline{i_n(t + \tau)}\}. \quad (2.1.8)$$

If we take the Fourier-transform of the correlation matrix of Equation (2.1.7), we will get the cross-spectral density matrix:

$$\mathbf{C}(\omega) = (c_{mn}(\omega)) = \left(\int_{-\infty}^{\infty} c_{mn}(\tau) \cdot e^{-j\omega\tau} \cdot d\tau \right). \quad (2.1.9)$$

Thus, for any arbitrary linear noisy network, we can reduce it to Equation (2.1.1) and Equation (2.1.4), with the statistical description of its noise behavior given by Equations (2.1.8) and (2.1.9).

2.2. Defining Vector Space for Linear Noisy Network

Based on the theory introduced in Section 2.1, classical noisy network modeling and analysis approach starts with reducing any given complex network into the compact general form. One application of this general form is to derive the minimum achievable noise figure for a general two-port noisy network by Adler and Haus [3]. In addition, one of the major applications of noisy network modeling is to describe the noise behavior of active devices, like transistors. Van der Ziel reduces the thermal noise contributed by the distributed resistors in a FET's channel to a two-port general form [7] [8].

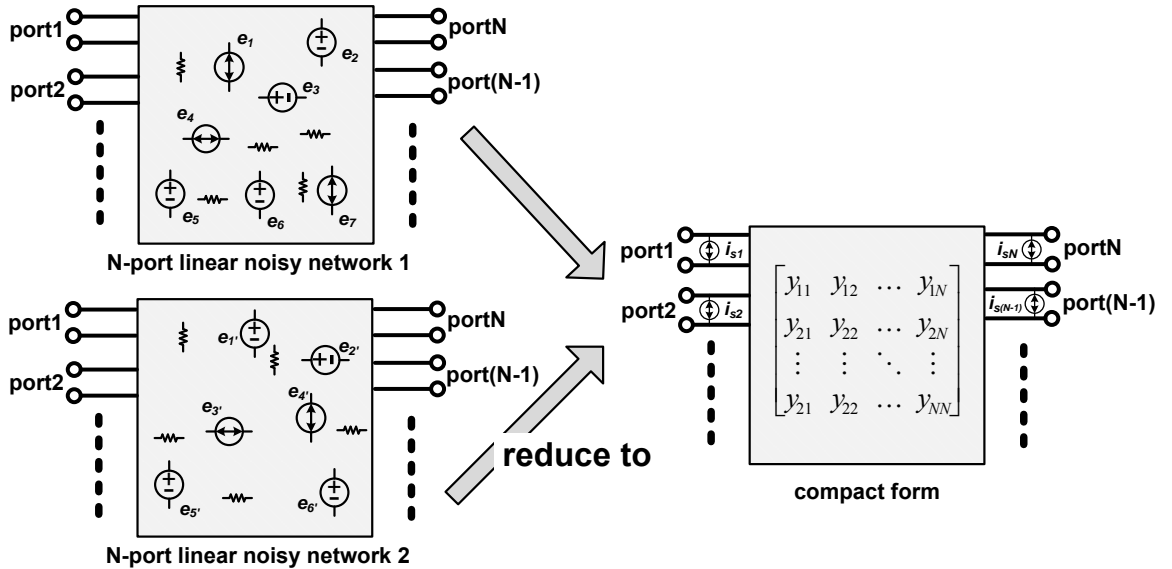


Figure 2.1: Different noisy networks might be able to reduce to the same compact network form.

Often in low-noise circuit designs, we will have to resort to EDA software to help us calculate the noise performance of a complex circuit system. Though reducing an elementary noisy network into a compact general form is neat in its mathematical expression, the correlation terms in Equations (2.1.7) and (2.1.9) between different noise sources are difficult to implement. What has been pointed out before is that it is possible for several different physical noisy networks to be reduced to a same general compact network (see example Figure 2.1). In other words, though these physical noisy networks may have different internal structures and noise sources, their network behaviors and statistical properties will be exactly the same when looking from the external world. Since different noisy network structures have different implementation difficulties, it makes it possible to choose to use the noisy network structure that is easiest to implement. However, we have to answer the problem: How do we find such a network in a systematic way? In order to answer this question, we have to look at the compact noisy network of Equation (2.1.4) from a different perspective.

The independent noise sources $i_{s1}(t), i_{s2}(t), \dots, i_{sN}(t)$ in Equation (2.1.4) are physical signals. They can be measured by connecting N ideal current meters to measure their short-circuit currents. This means that $i_{s1}(t), i_{s2}(t), \dots, i_{sN}(t)$ are real random processes. Since these random processes are real, their cross-correlation function will satisfy:

$$c_{mn}(\tau) = E\{i_m(t)\overline{i_n(t+\tau)}\} = E\{i_n(t+\tau)i_m(t)\} = E\{i_n(t)\overline{i_m(t-\tau)}\} = c_{nm}(-\tau).$$

Taking the Fourier transform of $c_{mn}(\tau)$, we will get $c_{mn}(\omega) = \overline{c_{nm}(\omega)}$. This means that

$c_{mn}(\omega)$ and $c_{nm}(\omega)$ are a complex conjugate pair. So the cross-correlation matrix will satisfy:

$$\mathbf{C}(\tau) = \mathbf{C}^T(-\tau). \quad (2.2.1)$$

And the cross-spectral density matrix satisfies:

$$\mathbf{C}(\omega) = \overline{\mathbf{C}^T(\omega)} \equiv \mathbf{C}^*(\omega). \quad (2.2.2)$$

The $[]^*$ operator takes the complex conjugate of the transpose of operand, and gives the adjoint matrix of the operand matrix [4]. In addition, the diagonal elements of the cross spectrum are real, since $c_{mm}(\omega) = \overline{c_{mm}(\omega)}$.

Based on these discussions, we can define $c_{mn}(\omega) \equiv r_{mn}(\omega) + jx_{mn}(\omega)$ for $m < n$, $x_{mn}(\omega) = 0$ for $m = n$, and $c_{mn}(\omega) = \overline{c_{nm}(\omega)} = r_{nm}(\omega) - jx_{nm}(\omega)$ for $m > n$. And $r_{mn}(\omega)$ and $x_{mn}(\omega)$ are real functions. So, the cross-spectral density matrix can be written as:

$$\mathbf{C}(\omega) = \begin{pmatrix} r_{11}(\omega) & r_{12}(\omega) + jx_{12}(\omega) & \dots & r_{1N}(\omega) + jx_{1N}(\omega) \\ r_{12}(\omega) - jx_{12}(\omega) & r_{22}(\omega) & \dots & r_{2N}(\omega) + jx_{2N}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ r_{1N}(\omega) - jx_{1N}(\omega) & r_{2N}(\omega) - jx_{2N}(\omega) & \dots & r_{NN}(\omega) \end{pmatrix} \quad (2.2.3)$$

At a given frequency ω , we can use N^2 real values to represent an N-port noisy network's noise behavior.

All physical noise sources inside an arbitrary N-port network are uncorrelated to each other internally. In Equation (2.1.4), $i_{s1}(t), i_{s2}(t), \dots, i_{sN}(t)$ has nonzero correlation because we are trying to model a complex internal network structure using a much simpler mathematical expression. Without the loss of generality, we assume that an arbitrary N-port

network has M uncorrelated noise sources, namely: $e_{n1}(t), e_{n2}(t), \dots, e_{nM}(t)$. And $E\{e_{ni}(t) \cdot e_{nj}(t)\} = 0$ for $i \neq j$. We can calculate the output currents $i_{s1}(t), i_{s2}(t), \dots, i_{sN}(t)$ as functions of these internal noise sources.

$$\begin{aligned} i_{1s}(t) &= h_{11}(t) * e_{n1}(t) + h_{12}(t) * e_{n2}(t) + \dots h_{1M}(t) * e_{nM}(t) \\ i_{2s}(t) &= h_{21}(t) * e_{n1}(t) + h_{22}(t) * e_{n2}(t) + \dots h_{2M}(t) * e_{nM}(t) \\ &\dots \\ i_{Ns}(t) &= h_{N1}(t) * e_{n1}(t) + h_{N2}(t) * e_{n2}(t) + \dots h_{NM}(t) * e_{nM}(t) \end{aligned} \quad (2.2.4)$$

Here, $h_{jk}(t)$ is the impulse response from the internal noise source e_{nk} to the output short-circuit current i_{js} with all ports shorted. $*$ is the convolution operator. The power spectral density of current sources: $i_{s1}(t), i_{s2}(t), \dots, i_{sN}(t)$ can be calculated to be:

$$\begin{aligned} S_{i_{js}, i_{js}}(\omega) &= |h_{j1}(\omega)|^2 \cdot S_{n1}(\omega) + |h_{j2}(\omega)|^2 \cdot S_{n2}(\omega) \\ &\dots + |h_{jM}(\omega)|^2 \cdot S_{nM}(\omega). \end{aligned} \quad (2.2.5)$$

Here, $S_{i_{js}, i_{js}}(\omega) \equiv \int_{-\infty}^{+\infty} E\{i_{js}(\tau) \cdot i_{js}(t - \tau)\} \cdot e^{-j\omega t} \cdot dt$, $S_{nm}(\omega) \equiv \int_{-\infty}^{+\infty} E\{e_{nm}(\tau) \cdot e_{nm}(t - \tau)\} \cdot e^{-j\omega t} \cdot dt$, and $h_{jk}(\omega) = \int_0^{+\infty} h_{jk}(t) \cdot e^{-j\omega t} \cdot dt$. We use the fact that all physical networks are causal. Similarly, we can calculate the cross-spectral density of the current sources:

$$\begin{aligned} S_{i_{js}, i_{qs}}(\omega) &= h_{j1}(\omega)h_{q1}^*(\omega) \cdot S_{n1}(\omega) + h_{j2}(\omega)h_{q2}^*(\omega) \cdot S_{n2}(\omega) + \\ &\dots + h_{jM}(\omega)h_{qM}^*(\omega) \cdot S_{nM}(\omega). \end{aligned} \quad (2.2.6)$$

Comparing Equations (2.2.3), (2.2.5), and (2.2.6), we realize that $r_{jj} = S_{i_{js}, i_{js}}(\omega)$, $r_{jq} = \text{Re}\{S_{i_{js}, i_{qs}}(\omega)\}$, and $x_{jq} = \text{Im}\{S_{i_{js}, i_{qs}}(\omega)\}$ for $j < q$. If we define the N^2 -tuples $(r_{11}, r_{12}, x_{12}, \dots, r_{1N}, x_{1N}, r_{22}, \dots, r_{2N}, x_{2N}, \dots, x_{NN})$ as a vector, the total noise behavior of

the N-port noisy network can be related to the magnitude of individual internal noise sources by:

$$\begin{bmatrix} r_{11}(\omega) \\ r_{12}(\omega) \\ x_{12}(\omega) \\ \vdots \\ r_{NN}(\omega) \end{bmatrix} = \begin{bmatrix} |h_{11}(\omega)|^2 & |h_{12}(\omega)|^2 & \dots & |h_{1M}(\omega)|^2 \\ \text{Re}\{h_{21}(\omega)h_{11}^*(\omega)\} & \text{Re}\{h_{22}(\omega)h_{12}^*(\omega)\} & \dots & \text{Re}\{h_{2M}(\omega)h_{1M}^*(\omega)\} \\ \text{Im}\{h_{21}(\omega)h_{11}^*(\omega)\} & \text{Im}\{h_{22}(\omega)h_{12}^*(\omega)\} & \dots & \text{Im}\{h_{2M}(\omega)h_{1M}^*(\omega)\} \\ \vdots & \vdots & \ddots & \vdots \\ |h_{N1}(\omega)|^2 & |h_{N2}(\omega)|^2 & \dots & |h_{NM}(\omega)|^2 \end{bmatrix}_{N^2 \times M} \begin{bmatrix} S_{n1}(\omega) \\ S_{n2}(\omega) \\ \vdots \\ S_{nM}(\omega) \end{bmatrix} \quad (2.2.7)$$

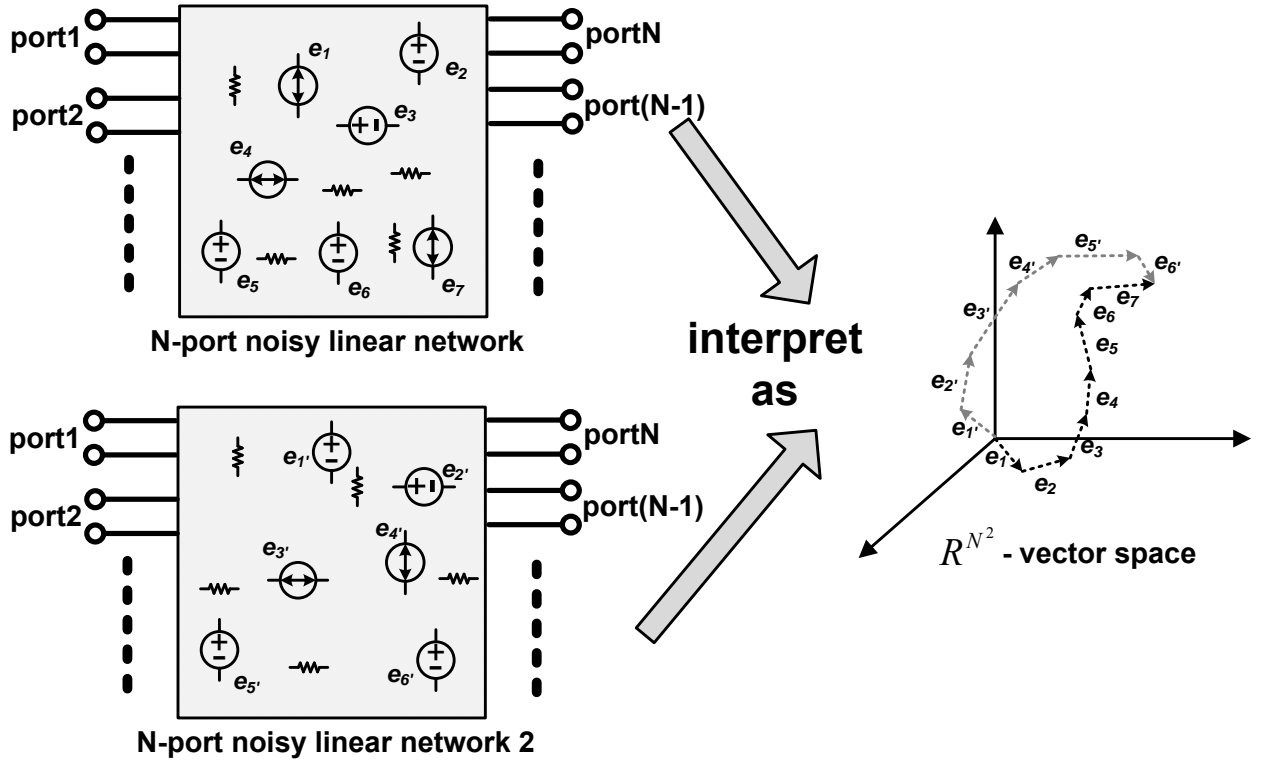


Figure 2.2: Noise contributions from the internal physical noise sources to the external world can be interpreted as the sum of several noise vectors in the defined vector space.

Now, if we define the N^2 -tuples $(r_{11}, r_{12}, x_{12}, \dots, r_{1N}, x_{1N}, r_{22}, \dots, r_{2N}, x_{2N}, \dots, x_{NN})$ as R^{N^2} -vector space, we can interpret the noise process in the N-port noisy network such that each internal noisy source e_{nk} contributes a small vector in the defined R^{N^2} -vector space. And the total noise behavior of the N-port noisy network is the vector sum of these small vectors contributed by all internal noise sources. In Figure 2.2, we use an example of N-port network with seven internal physical noise sources to demonstrate the concept.

There are several implications of interpreting an arbitrary noisy network in this manner. First of all, if two noisy networks with different internal noise sources accumulate to a same-summed noisy vector, their statistical behavior would be the same from the external world. As shown in Figure 2.2, network 1 and network 2 have different internal structures, and different number of noise sources. The contribution of these noise sources inside the two noisy networks will correspond to two different “trajectories” in the defined R^{N^2} -vector space. However, since their vector sums point to the same point in the vector space, network 1 and network 2 have the same statistical behavior.

Now, since a R^{N^2} -vector space can be used to interpret an arbitrary N-port noisy network, if we can find a set of N^2 noise sources, which are uncorrelated with each other and are linearly independent in the R^{N^2} -vector space, we can completely model an N-port noisy network. The requirement of N^2 uncorrelated noise sources is the worst case scenario. If the rank of the noisy network is smaller than N^2 , some of these noise sources are unnecessary.

One final remark before the end of this section: There are several different network representations of an N-port linear noisy network, and in this section, we choose the admittance matrix representations and define the R^{N^2} -vector space based on it. If we choose a different network representation, say the impedance matrix, we will get a different R^{N^2} -vector space. However, they are mathematically equivalent and can be converted to one another by a linear transformation.

In the next section, we will use this concept to model a two-port noisy network as a general two-port example.

2.3. Example: A Two-Port Noisy Network

Classical approach of two-port noise modeling reduces a given noisy network into Equation (2.1.1). Due to the correlation between the two elements in $\mathbf{I}_s = [i_{s1}(s) \ i_{s2}(s)]^T$, special efforts need to be taken in order to simulate an arbitrary noisy two-port network. One possible method is to utilize two correlation admittances at the input port to decorrelate the two noisy sources [9], as shown in Figure 2.3. The overhead of this approach is the necessity of constructing an “embedding” network.

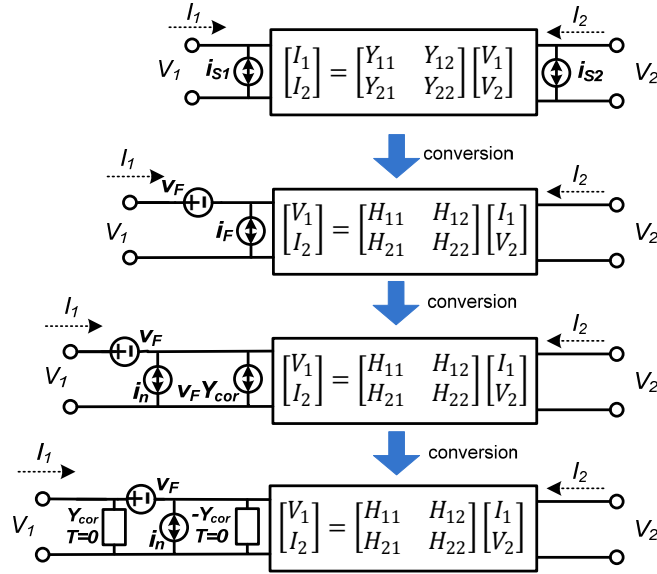


Figure 2.3: Two correlation admittances are used to decorrelate the two noisy sources.

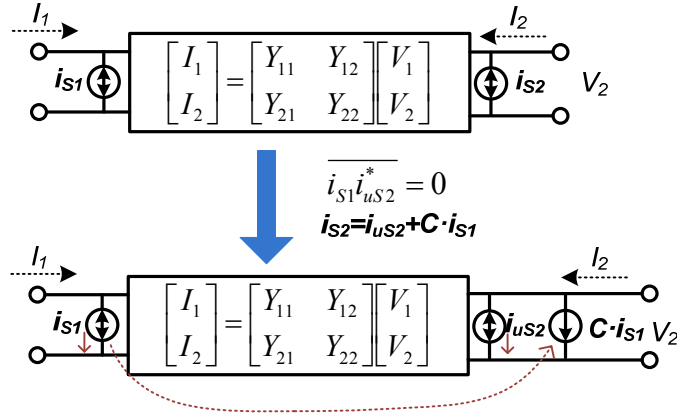


Figure 2.4: A controlled source is used to implement a noisy two-port network.

Another commonly used approach is to separate the second noise source (i_{S2}) into a part that is fully correlated with the first noise source (i_{S1}) and an other part (i_{uS2}) that is totally uncorrelated with i_{S1} , as shown in Figure 2.4. A controlled source is used to introduce the correlation between two correlated noise sources (i_{S1} and $C \cdot i_{S1}$).

We can also use the concept introduced in Section 2.2 to model an arbitrary noisy network with $N^2 = 4$ ($N = 2$) noise sources. As shown in Figure 2.5(a), we have an arbitrary physical two-port linear noisy network, with some arbitrary internal circuit connections and physical noise sources. The classical approach reduces the given network into a compact form shown in Figure 2.5(b) as a basis for circuit analysis. If we define a R^4 -vectors space by grouping $[S_{i_{S_1}}(\omega) \quad Re\{S_{i_{S_1}\bar{i}_{S_2}}(\omega)\} \quad Im\{S_{i_{S_1}\bar{i}_{S_2}}(\omega)\} \quad S_{i_{S_2}}(\omega)]^T$, we can plot the contributions of the internal noise sources in Figure 2.5(a) in the R^4 -vectors space as several small noisy vectors. The overall statistical behavior of the given arbitrary network is thus a vector sum of these smaller noisy vectors, as shown in Figure 2.5(d). It should be noted that, for the convenience of plotting the concept, we use five internal noisy sources for the network in Figure 2.5(a). In general, the number of noise sources inside the noisy network can be arbitrary.

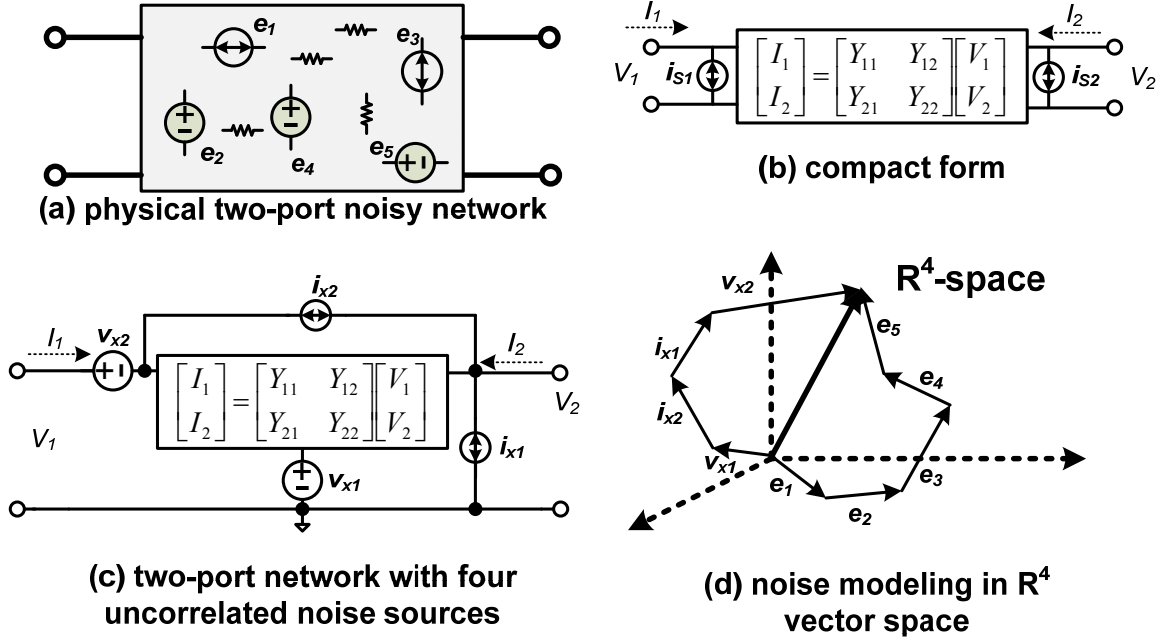


Figure 2.5: (a) A physical two-port linear noisy network, (b) Compact form of a two-port noisy network, (c) A two-port network with four uncorrelated noise sources, and (d) The conceptual plots of the noise vectors for (a) and (c) in an R^4 -space

Since any point in the defined R^4 vector space represents a particular statistical behavior, we can find another noisy network with four uncorrelated noise sources to match an arbitrary two-port network's noise property. In Figure 2.5(c), we show one of the possible network choices. We choose this network topology for the convenience of modeling an FET. In general, we can choose arbitrary four-noise sources as long they are linearly independent in the R^4 -space. To model an arbitrary two-port network with the network in Figure 2.5(c), we need to first relate i_{s1} and i_{s2} in Figure 2.5(a) by v_{x1} , v_{x2} , i_{x1} , and i_{x2} :

$$\begin{aligned}
i_{S1} &= -(y_{11} + y_{12}) \cdot v_{x1} - y_{11} \cdot v_{x2} + i_{x2} \\
i_{S2} &= -(y_{21} + y_{22}) \cdot v_{x1} - y_{21} \cdot v_{x2} + i_{x1} - i_{x2}.
\end{aligned} \tag{2.3.1}$$

Based on Equation (2.3.1), we can calculate the spectral density and the cross spectral density of i_{S1} and i_{S2} in terms of $S_{v_{x1}}$, $S_{v_{x2}}$, $S_{i_{x1}}$, and $S_{i_{x2}}$.

$$\begin{aligned}
S_{i_{S1}} &= |y_{11} + y_{12}|^2 \cdot S_{v_{x1}} + |y_{11}|^2 \cdot S_{v_{x2}} + S_{i_{x2}} \\
S_{i_{S2}} &= |y_{21} + y_{22}|^2 \cdot S_{v_{x1}} + |y_{21}|^2 \cdot S_{v_{x2}} + S_{i_{x1}} + S_{i_{x2}} \\
S_{i_{S1}\overline{i_{S2}}} &= (y_{11} + y_{12}) \cdot \overline{(y_{21} + y_{22})} \cdot S_{v_{x1}} + y_{11} \cdot \overline{y_{21}} \cdot S_{v_{x2}} - S_{i_{x2}}
\end{aligned} \tag{2.3.2}$$

Grouping $[S_{i_{S1}}(\omega) \quad \text{Re}\{S_{i_{S1}\overline{i_{S2}}}(\omega)\} \quad \text{Im}\{S_{i_{S1}\overline{i_{S2}}}(\omega)\} \quad S_{i_{S2}}(\omega)]^T$ into an R^4 space, we can rewrite Equation (2.3.2) as:

$$\begin{aligned}
&\begin{bmatrix} S_{i_{S1}} \\ \text{Re}\{S_{i_{S1}\overline{i_{S2}}}\} \\ \text{Im}\{S_{i_{S1}\overline{i_{S2}}}\} \\ S_{i_{S2}} \end{bmatrix} = \\
&\begin{bmatrix} |y_{11} + y_{12}|^2 & |y_{11}|^2 & 0 & 1 \\ \text{Re}\{(y_{11} + y_{12}) \cdot \overline{(y_{21} + y_{22})}\} & \text{Re}\{y_{11} \cdot \overline{y_{21}}\} & 0 & -1 \\ \text{Im}\{(y_{11} + y_{12}) \cdot \overline{(y_{21} + y_{22})}\} & \text{Im}\{y_{11} \cdot \overline{y_{21}}\} & 0 & 0 \\ |y_{22} + y_{21}|^2 & |y_{21}|^2 & 1 & 1 \end{bmatrix} \begin{bmatrix} S_{v_{x1}} \\ S_{v_{x2}} \\ S_{i_{x1}} \\ S_{i_{x2}} \end{bmatrix}.
\end{aligned} \tag{2.3.3}$$

The criteria for the network in Figure 2.5(c) to have a solution is that the linearly independent condition needs to be hold. Linearly independent condition can hold if and only if:

$$\det \begin{bmatrix} |y_{11} + y_{12}|^2 & |y_{11}|^2 & 0 & 1 \\ \text{Re}\{(y_{11} + y_{12}) \cdot \overline{(y_{22} + y_{21})}\} & \text{Re}\{y_{11} \cdot \overline{y_{21}}\} & 0 & -1 \\ \text{Im}\{(y_{11} + y_{12}) \cdot \overline{(y_{22} + y_{21})}\} & \text{Im}\{y_{11} \cdot \overline{y_{21}}\} & 0 & 0 \\ |y_{22} + y_{21}|^2 & |y_{21}|^2 & 1 & 1 \end{bmatrix} \neq 0. \tag{2.3.4}$$

In the next section, we will use this two-port noisy network example of Figure 2.5(c) to match a noisy intrinsic FET, based on Van der Ziel's model.

2.4. A Modified FET Noise Model

Van der Ziel attributes the noise of an intrinsic FET to the distributed resistors inside the channel of a FET. As summarized in Appendix 2.1 of this chapter, he reduced the aggregate distributed thermal noise into a drain thermal noise (i_{s1}) and an induced gate noise (i_{s2}). Due to these two noises being generated from the same physical distributed resistors inside the channel, the drain noise and the gate noise are correlated. His derivations show:

$$S_{i_{s1}}(\omega) = \frac{4kT[\omega \cdot C_{gs}]^2}{5g_{d0}} \cdot \delta \quad (2.5.1)$$

$$S_{i_{s2}}(\omega) = 4kTg_{d0} \cdot \gamma \quad (2.5.2)$$

$$S_{i_{s1}i_{s2}}(\omega) = j|c| \cdot \sqrt{S_{i_{s1}}(\omega) \cdot S_{i_{s2}}(\omega)}. \quad (2.5.3)$$

For a long-channel FET, $\gamma = \frac{2}{3}$, $\delta = \frac{4}{3}$, and coefficient $c = \frac{1}{6} \cdot \sqrt{\frac{45}{8}} \approx 0.395$. For an intrinsic FET, its admittance matrix can also be found to be:

$$Y(\omega) = \begin{bmatrix} j\omega \cdot C_{gs} & 0 \\ g_m & g_{ds} \end{bmatrix}. \quad (2.5.4)$$

Note that, in Van der Ziel's original derivations, the gate-to-drain capacitance C_{gd} is extrinsic.

Based on the general two-port network example in Section 2.3, we can use four uncorrelated noise sources to model the Van der Ziel's derived intrinsic FET model by solving:

$$\begin{bmatrix} \frac{4kT[\omega \cdot C_{gs}]^2}{5g_{d0}} \cdot \delta \\ 0 \\ 4kT|c|\omega C_{gs} \cdot \sqrt{\frac{\delta\gamma}{5}} \\ 4kTg_{d0} \cdot \gamma \end{bmatrix} = \quad (2.5.5)$$

$$\begin{bmatrix} \omega^2 C_{gs}^2 & \omega^2 C_{gs}^2 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ \omega C_{gs}(g_m + g_{ds}) & \omega C_{gs}g_m & 0 & 0 \\ g_m^2 & g_m^2 & 1 & 1 \end{bmatrix} \begin{bmatrix} S_{v_{xs}} \\ S_{v_{xg}} \\ S_{i_{xd}} \\ S_{i_{xgd}} \end{bmatrix}$$

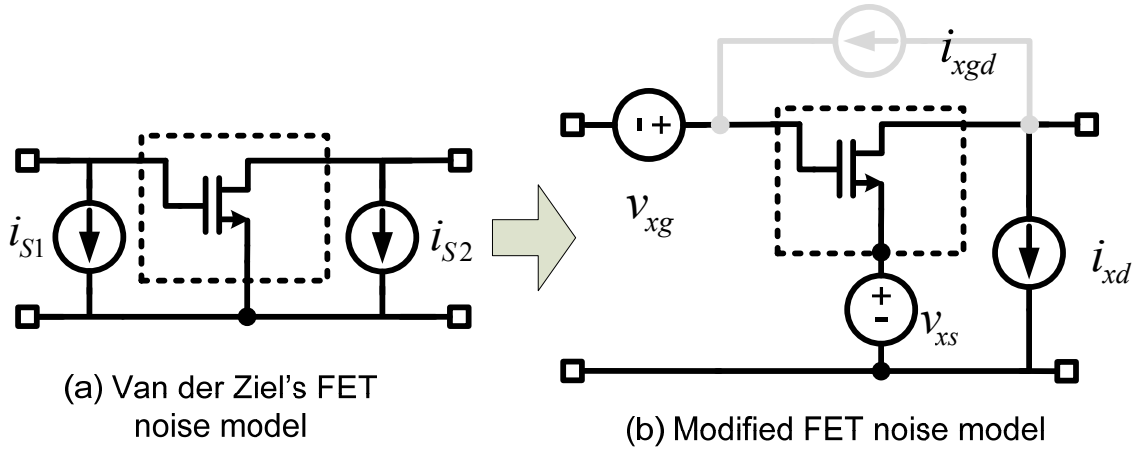


Figure 2.6: Three frequency-independent and uncorrelated sources are used to implement Van der Ziel's FET noise model.

This process is shown in Figure 2.6, and the solution of the above linear equations is:

$$\begin{bmatrix} S_{v_{xs}} \\ S_{v_{xg}} \\ S_{i_{xd}} \\ S_{i_{xgd}} \end{bmatrix} = 4kT \begin{bmatrix} \frac{|c|}{g_{ds}} \sqrt{\frac{\delta\gamma}{5}} - \frac{g_m\delta}{5g_{d0}g_{ds}} \\ \frac{\delta}{5g_{d0}} - \frac{|c|}{g_{ds}} \sqrt{\frac{\delta\gamma}{5}} + \frac{g_m\delta}{5g_{d0}g_{ds}} \\ g_{d0}\gamma - \frac{\delta g_m^2}{5g_{d0}} \\ 0 \end{bmatrix}. \quad (2.5.6)$$

There are several interesting characteristics of this solution. First of all, $S_{i_{xgd}} = 0$, which means that we will only need three uncorrelated noise sources instead of four to

implement Van der Ziel's FET noise model. In addition, the nonzero noise sources: $S_{v_{xg}}$, $S_{v_{xs}}$, and $S_{v_{xd}}$ are frequency independent, so they can be implemented using two white noise voltage sources and a white noise current source. Since both white noise voltage source and white current voltage source are supported by almost all EDA tools, the modified FET model in Figure 2.6 can be easily implemented in an EDA design environment. Furthermore, in the modified FET model, all three noise sources are uncorrelated with each other, this will make the hand calculation of a complex noisy network consisting of many FET transistors much easier.

As mentioned in Section 2.3, we can choose any four uncorrelated noise sources to model an arbitrary noisy two-port network, as long as these four noise sources are linearly independent in the R^4 -vector space. Figure 2.7 shows another modified FET noise model with four different noise sources: i_{1x} , i_{2x} , i_{3x} , and v_x . Matching the Van der Ziel's model with the noise model in Figure 2.7 (b), we will get the spectral density of these noise sources:

$$\begin{bmatrix} S_{v_x} \\ S_{i_{1x}} \\ S_{i_{2x}} \\ S_{i_{3x}} \end{bmatrix} = 4kT \begin{bmatrix} \sqrt{\frac{\delta\gamma}{5}} \cdot \left(\frac{|c|}{g_m + g_{ds}} \right) \\ (\omega C_{gs})^2 \cdot \left\{ \frac{\delta}{5g_{d0}} - \sqrt{\frac{\delta\gamma}{5}} \cdot \left(\frac{|c|}{g_m + g_{ds}} \right) \right\} \\ g_{d0}\gamma - (g_m + g_{ds}) \cdot |c| \cdot \sqrt{\frac{\delta\gamma}{5}} \\ 0 \end{bmatrix}. \quad (2.5.7)$$

This solution of $S_{i_{1x}}$ has a frequency-dependent spectral density, but the overall solution is less sensitive to the value of g_{ds} , as compared with the network in Figure 2.6.

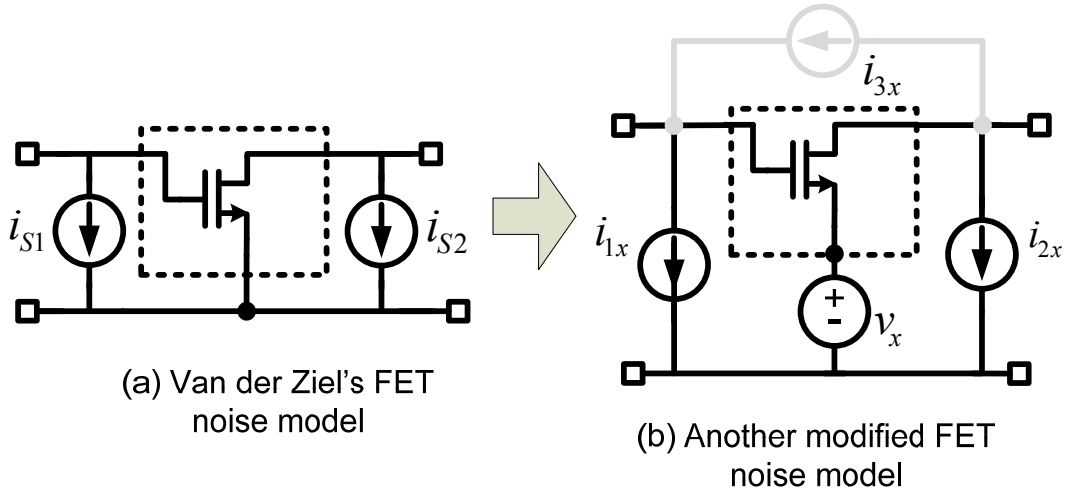


Figure 2.7: Another modified FET noise model is also equivalent to Van der Ziel's FET noise model.

2.5. FET Noise Model Comparisons

In addition to Van der Ziel's FET noise model, Pospieszalski's [10][11] model and the holistic noise model in the Berkeley Short-channel IGFET Model-version 4 (BSIM4) [12] are two other commonly used FET noise models. When modeling FET's noise behavior, confusions between the underlying physical causes of the thermal noise, and the mathematical completeness of a given model in the R^4 -vector space should be clarified.

Model Name	Physical Explanation	Mathematical Completeness
Van der Ziel	Noise from intrinsic FET is due to the distributed resistors in the channel.	Use general admittance matrix with two correlated sources. Mathematically complete.
Pospieszalski	Noise of FET is generated by	Use two uncorrelated sources.

	drain conductance at temperature T_G , and the gate resistance at T_D .	Mathematically incomplete.
BSIM4	N/A	Use two uncorrelated sources. Mathematically incomplete.
Modified FET model with three uncorrelated sources	N/A	Use three uncorrelated sources. Mathematically incomplete.
General two-port noisy model with four uncorrelated sources	N/A	Use four uncorrelated sources. Mathematically complete.

Table 2.1: Comparisons between FET noise models

On the one hand, both Van der Ziel's and Pospieszalski's models postulate the physical noise generation process inside a FET. A good physical noise model relates the structural parameters of a FET to its measurable noise behavior, and in the ideal scenario, the theory should match the measurement. On the other hand, we need to fit the noise measurement of a FET by our models, and a particular FET model may not have enough mathematical completeness to match the measurement. In other words, if the theory of noise process inside a FET deviates from what the real-world situation is, a particular noise model will not be able to fit it. If the noisy network representation has a degree of freedom that is less than four, it may not be enough to match the measurement, since a noise parameter of a noisy network has a dimension of four. In measurement, the necessity of de-embedding the parasitic networks from the intrinsic FET in microwave frequencies and the random nature of a noise measurement further complicate the modeling process. We summarize the mathematical completeness of commonly used FET models in Table 2.1.

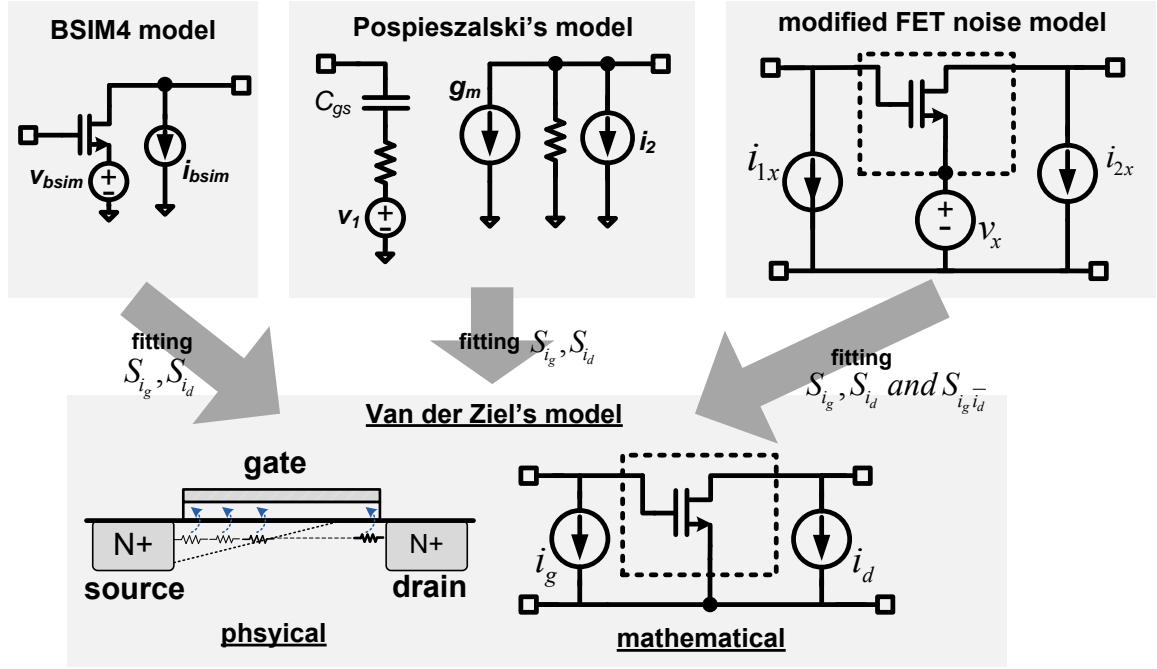


Figure 2.8: Pospieszalski's model, BSIM4 model, and modified FET noise model are fitted to the long-channel Van der Ziel's model.

Another way to compare different FET models is to directly fit models to experimental results [13]. In this approach, the FET noise measurement itself may not be representative. Instead of fitting a particular FET measurement, we will do a mutual fitting between a chosen physical model and the rest of the FET models. The purpose of this comparison is to illustrate how insufficient degree of freedom might affect the noise modeling, but not to argue the correctness of a physical model. We would first assume we have a FET device which follows Van der Ziel's long channel FET noise model with uniform channel temperature equal to ambient temperature T . We then fit Pospieszalski's model, the BSIM4 model, and the modified FET models with the FET's S_{i_g} , S_{i_d} , and $S_{i_g \bar{i}_d}$. Both

Pospiezalski's and the BSIM4 model have a dimension of two, so we will fit S_{i_g} and S_{i_d} and leave $S_{i_g\overline{d}}$ as a dependent variable. The model fitting process is shown in Figure 2.8.

The fitting results are plotted on the normalized $\left[\frac{S_{i_g}}{S_{i_g,vdz}}, \frac{im\{S_{i_g\overline{d}}\}}{im\{S_{i_g\overline{d}}\}_{vdz}}, \frac{S_{i_d}}{S_{i_d,vdz}} \right]^T$ vector space,

as the noise vector summation from the internal uncorrelated noise sources. Here, $S_{i_g,vdz} =$

$$\frac{4kT[\omega \cdot C_{gs}]^2 \cdot \delta}{5g_{d0}}$$

$$S_{i_d,vdz} = 4kTg_{d0} \cdot \gamma, \text{ and}$$

$$im\{S_{i_g\overline{d}}\}_{vdz} = |c| \cdot \sqrt{S_{i_{s1}}(\omega) \cdot S_{i_{s2}}(\omega)}.$$

$Re\{S_{i_g\overline{d}}\}$ is omitted because it is zero in Van der Ziel's model. The Van der Ziel's noise model itself is plotted as a trajectory from the origin to reflect the fact that the FET's noise is an aggregate behavior of the thermal noise generated from the distributed resistors inside the channel of a FET. Comparison results are plotted in Figure 2.9.

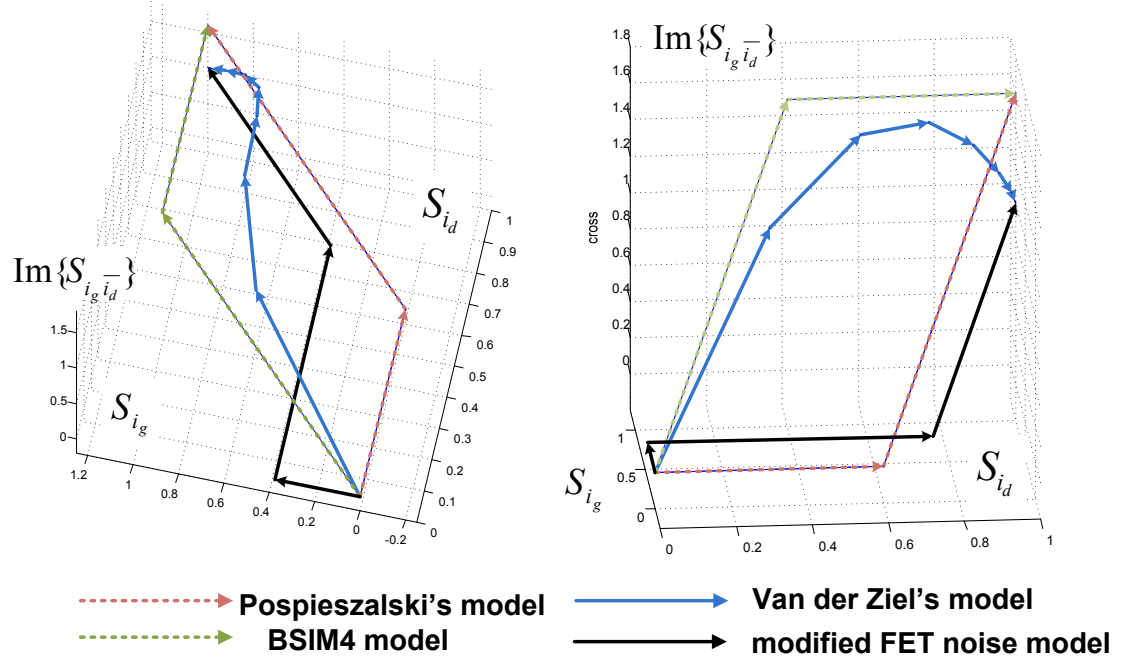


Figure 2.9: FET noise modeling is viewed as vector summation in the normalized vector space.

From this comparison plot, we understand that the modified FET noise model matches Van der Ziel's noise model, while both the BSIM4 and Pospieszalski's model leave errors in $S_{i_g i_d}$. This comparison agrees with the conclusions in [13].

2.6. Summary

In this chapter, we define a R^{N^2} -vector space for an arbitrary noisy network, and prove that any internal physical sources inside the noisy network contribute a small vector in the defined R^{N^2} -vector space, and the aggregate statistical behavior of this noisy network can be viewed as the vector sum of these small vectors.

A general two-port noisy network is demonstrated as an example. Its application to modeling the FET leads to a modified noise model of the FET, in which three uncorrelated noise sources are sufficient to describe the statistical behavior of an intrinsic FET. Comparisons between the modified noise model and existing models show that our new model fits Van der Ziel's model better than the others.

Appendix 2.1: Van der Ziel's FET Noise Model

Van der Ziel attributes the thermal noise of a FET to the distributed resistors inside the channel of an FET [7][8]. He also assumes quasi-stationary and a zero-order approximation of a noise perturbation inside the channel to simplify the calculation. These conditions are satisfied for normal FET operation, and Shoji [14] discusses when these conditions do not hold.

Now, if zero-order approximation inside a FET's channel is assumed, a small perturbation due to the thermal noise generated by the distributed resistor at location x_0 will give rise to a linear voltage perturbation distribution $\Delta V(x)$ along the channel on top of the DC equilibrium voltage $V_0(x)$. This $\Delta V(x)$ distribution is plotted in Figure 2.10.

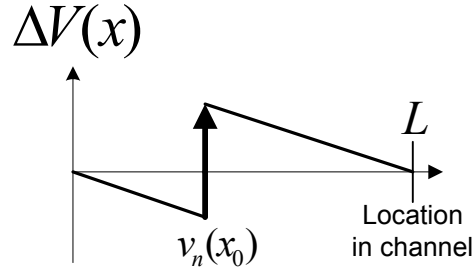


Figure 2.10: Linear voltage perturbation distribution along the channel of a FET due to a noise perturbation at x_0

Since the drain current of an FET is generated from the voltage gradient $\frac{\partial V_0}{\partial x}$ along the channel, any perturbation in this voltage will generate a corresponding drain current perturbation i_d . We can relate $v_n(x_0)$ to i_d by solving the partial differential equations:

$$\begin{aligned}\frac{\partial [G(V_0(x))\Delta V(x)]}{\partial x} &= i_d(t) \\ \Delta V(0) &= 0 \\ \Delta V(L) &= 0\end{aligned}\tag{2.7.1}$$

$$\Delta V(x_0 + dx) = \Delta V(x_0) + v_n(x_0, t).$$

And we will get:

$$\begin{aligned}\Delta V(x) &= -\frac{x \cdot G(V_0(x_0))}{L \cdot G(V_0(x))} \cdot v_n(x_0, t) \text{ for } 0 < x < x_0 \\ \Delta V(x) &= -\frac{(x-L) \cdot G(V_0(x_0))}{L \cdot G(V_0(x))} \cdot v_n(x_0, t) \text{ for } x_0 < x < L \\ i_d(x_0, \Delta x, t) &= -\frac{G(V_0(x_0))}{L} \cdot v_n(x_0, t).\end{aligned}\tag{2.7.2}$$

Note that this noise current perturbation $i_d(t)$ is due to the resistance between $(x_0, x_0 + \Delta x)$; we rewrite the $i_d(t)$ in Equation (2.7.1) as $i_d(x_0, \Delta x, t)$ in Equation (2.7.2). We also rewrite the r.m.s. value of $v_n(x_0, t)$ as $E\{|v_n(x_0, \Delta x, t)|^2\} = \frac{4kT\Delta F \cdot \Delta x}{G(V_0(x_0))}$. Also note that, $v_n(x_0, t)$ is white noise, and $v_n(x_0, t)$ is uncorrelated with $v_n(x_1, t)$ for $x_0 \neq x_1$. So:

$$\begin{aligned}E\{v_n(x_0, \Delta x, t + \tau) \cdot \overline{v_n(x_0, \Delta x, t)}\} &= \frac{4kT\Delta F \cdot \Delta x \cdot \delta(\tau)}{G(V_0)} \\ E\{v_n(x_0, \Delta x, t + \tau) \cdot \overline{v_n(x_1, \Delta x, t)}\} &= 0, \text{ for } x_0 \neq x_1.\end{aligned}\tag{2.7.3}$$

Drain noise due to resistance between $(x_0, x_0 + \Delta x)$

In a long channel quasi-static model of FET, $G(V_0(x)) \frac{dV_0(x)}{dx} = I_0$ is satisfied along the channel for $x \in [0, L]$. We also know that $G(V_0(x)) = \frac{G_{d0}}{V_{GS} - V_{TH}} \cdot (V_{GS} - V_0(x) - V_{TH})$ where $G_{d0} \equiv C_{ox} W \mu (V_{GS} - V_{TH})$. To get the drain noise due to the noise voltage $v_n(x_0, t)$, we take the auto-correlation of Equation (2.7.1), and we will get:

$$\begin{aligned} R_{i_d i_d}(x_0, \Delta x, \tau) &= E\{i_d(x_0, \Delta x, t) \cdot \overline{i_d(x_0, \Delta x, t + \tau)}\} \\ &= \frac{4kT\Delta F \cdot G(V_0(x_0))}{L^2} \delta(\tau) \cdot \Delta x. \end{aligned} \quad (2.7.4)$$

The power spectral density of $i_d(x_0, \Delta x)$ is thus:

$$S_{i_d i_d}(x_0, \Delta x) = \frac{4kT\Delta F \cdot G(V_0(x_0))}{L^2} \cdot \Delta x. \quad (2.7.5)$$

If we take the limit of $\Delta x \rightarrow 0$ and simplify the equation using the quasi-static assumption, we will get:

$$\lim_{\Delta x \rightarrow 0} \{S_{i_d i_d}(x_0, \Delta x)\} = \frac{4kT\Delta F \cdot G^2(V_0(x_0))}{L^2 I_0} \cdot dV_0(x_0). \quad (2.7.6)$$

Gate noise due to resistance between $(x_0, x_0 + \Delta x)$

As shown in Figure 2.10, a voltage perturbation at x_0 generates a voltage perturbation distribution $\Delta V(x)_{due\ x_0}$ along the channel. This voltage distribution will need to be accompanied by the charge distribution $\Delta Q_g(x)$ on the other (gate) side of the MOS structure, and they are related by:

$$\Delta Q_g(x) = -C_{ox} \cdot W \cdot \Delta V(x)_{due\ x_0}. \quad (2.7.7)$$

The total charge accumulation due to the noise voltage $v_n(x_0, \Delta x)$ can be found by integrating Equation (2.7.7) along x , using the results of Equation (2.7.2), and we get:

$$\Delta Q_g(x_0, \Delta x, t) = \frac{C_{ox}^2 W^2 \cdot \mu \cdot G(V_0(x_0))}{2LI_0^2} \cdot \left\{ V_{DS}^2 \cdot (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox}W\mu} (V_0(x_0) - V_{DS}) \right\} \cdot v_n(x_0, \Delta x, t). \quad (2.7.8)$$

Here, $\Delta Q_g(x_0, \Delta x, t)$ is the total charge accumulation due to a noise voltage $v_n(x_0, \Delta x, t)$ at x_0 . Since $v_n(x_0, \Delta x, t)$ changes over time, so does $\Delta Q_g(x_0, \Delta x, t)$, if we short the gate of an FET, we will observe a time-varying current $i_g(x_0, \Delta x, t)$ related to the $\Delta Q_g(x_0, \Delta x, t)$ by: $i_g(x_0, \Delta x, t) = \frac{\partial \Delta Q_g(x_0, \Delta x, t)}{\partial t}$. The direction of the $i_g(x_0, \Delta x, t)$ is also important when calculating correlation between i_g and i_d , and this is shown in Figure 2.11. Note that the choice of the direction of i_g in Figure 2.11 is opposite to that in Van der Ziel's original paper.

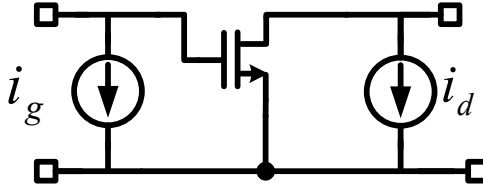


Figure 2.11: Direction of the current sources

Differentiate Equation (2.7.8) with time, and we will get:

$$i_g(x_0, \Delta x, t) = \frac{C_{ox}^2 W^2 \cdot \mu \cdot G(V_0(x_0))}{2LI_0^2} \cdot \left\{ V_{DS}^2 \cdot (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox}W\mu} (V_0(x_0) - V_{DS}) \right\} \cdot \frac{\partial v_n(x_0, \Delta x, t)}{\partial t}. \quad (2.7.9)$$

Taking the autocorrelation of Equation (2.7.9), we will get:

$$R_{i_g i_g}(x_0, \Delta x, \tau) = \frac{C_{ox}^4 W^4 \mu^2 G^2 (V_0(x_0))}{4L^2 I_0^4} \cdot \left\{ V_{DS}^2 \cdot (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox} W \mu} (V_0(x_0) - V_{DS}) \right\}^2 \cdot E \left\{ \frac{\partial v_n(x_0, \Delta x, t)}{\partial t} \cdot \frac{\partial v_n(x_0, \Delta x, t + \tau)}{\partial t} \right\}. \quad (2.7.10)$$

So, the spectral density of i_g will become:

$$\lim_{\Delta x \rightarrow 0} S_{i_g i_g}(x_0, \Delta x, \omega) = \frac{kT \omega^2 C_{ox}^4 W^4 \mu^2 G^2 (V_0(x_0))}{L^2 I_0^5} \cdot \left\{ V_{DS}^2 \cdot (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox} W \mu} (V_0(x_0) - V_{DS}) \right\}^2 \cdot dV_0(x_0). \quad (2.7.11)$$

We use the fact that $\int_{-\infty}^{+\infty} E \left\{ \frac{\partial v_n(x_0, \Delta x, t)}{\partial t} \cdot \frac{\partial v_n(x_0, \Delta x, t + \tau)}{\partial t} \right\} = \omega^2 \cdot S_{v_n v_n}(\omega)$.

Cross-correlation of gate noise and drain noise due to resistance between $(x_0, x_0 + \Delta x)$

Taking the cross-correlation of i_g and i_d is defined as:

$$R_{i_g i_d}(x_0, \Delta x, \tau) = E \{ i_g(t + \tau) \cdot i_d(t) \} = - \frac{C_{ox}^2 W^2 \cdot \mu \cdot G^2 (V_0(x_0))}{2L^2 I_0^2} \cdot \left\{ V_{DS}^2 \cdot (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox} W \mu} (V_0(x_0) - V_{DS}) \right\} \cdot E \left\{ \frac{\partial v_n(x_0, \Delta x, t + \tau)}{\partial t} \cdot v_n(x_0, t) \right\}. \quad (2.7.12)$$

So their cross-spectral density will be:

$$\lim_{\Delta x \rightarrow 0} S_{i_g i_d}(x_0, \Delta x, \tau) = -j\omega \cdot \frac{4kTC_{ox}^2 W^2 \cdot \mu \cdot G^2 (V_0(x_0))}{2L^2 I_0^3} \cdot \left\{ V_{DS}^2 \cdot (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox} W \mu} (V_0(x_0) - V_{DS}) \right\} \cdot dV_0(x_0). \quad (2.7.13)$$

Vectored contribution due to resistance between $(x_0, x_0 + \Delta x)$

Summarizing Equations (2.7.6), (2.7.11), and (2.7.13), a distributed resistor between

$\lim_{\Delta x \rightarrow 0}(x_0, x_0 + \Delta x)$ in the channel, will contribute:

$$\left[\begin{aligned} & \frac{kT\omega^2 C_{ox}^4 W^4 \mu^2 G^2 (V_0(x_0))}{L^2 I_0^5} \left\{ V_{DS}^2 (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox}W\mu} (V_0(x_0) - V_{DS}) \right\}^2 dV_0(x_0) \\ & - \frac{2kT\omega C_{ox}^2 W^2 \mu G^2 (V_0(x_0))}{L^2 I_0^3} \left\{ V_{DS}^2 (V_{GS} - V_{TH}) - \frac{1}{3} V_{DS}^3 + \frac{2LI_0}{C_{ox}W\mu} (V_0(x_0) - V_{DS}) \right\} dV_0(x_0) \\ & \frac{4kT \cdot G^2 (V_0(x_0))}{L^2 I_0} dV_0(x_0) \end{aligned} \right] \quad (2.7.14)$$

in the $[S_{igig}, im\{S_{igid}\}, S_{idid}]^T$ vector space.

Total noise contribution due to resistance between $(0, x_0)$

Simplifying Equation (2.7.14) with change of variables:

$$\begin{aligned} \eta(x_0) &\equiv \frac{V_0(x_0)}{V_{GS} - V_{TH}} \\ d\eta(x_0) &= \frac{dV_0(x_0)}{V_{GS} - V_{TH}} \\ \eta(L) &= \frac{V_0(L)}{V_{GS} - V_{TH}} = \frac{V_{DS}}{V_{GS} - V_{TH}}, \end{aligned} \quad (2.7.15)$$

then:

$$\begin{aligned} G(V_0(x_0)) &= C_{ox}W\mu \cdot [V_{GS} - V_0(x_0) - V_{TH}] \\ &= C_{ox}W\mu \cdot (V_{GS} - V_{TH}) \cdot [1 - \eta(x_0)] \end{aligned} \quad (2.7.16)$$

and

$$I_0 = \frac{C_{ox}W\mu(V_{GS} - V_{TH})^2}{L} \cdot \left[1 - \frac{1}{2}\eta(L) \right] \cdot \eta(L). \quad (2.7.17)$$

We then integrate Equation (2.7.14) from $x = 0$ to x_0 , and we will get:

$$\begin{aligned}
S_{i_g i_g} \Big|_{(0, x_0)} &= \frac{kT \omega^2 C_{ox}^2 W^2 L^2}{g_{d0} \cdot \left[1 - \frac{1}{2} \eta(L)\right]^5 \cdot \eta^5(L)} \\
&\cdot \left\{ \frac{1}{5} A^2 \eta^5 + \frac{1}{4} (2AB - 2A^2) \eta^4 + \frac{1}{3} (A^2 + B^2 - 4AB) \eta^3 \right. \\
&\quad \left. + \frac{1}{2} (2AB - 2B^2) \eta^2 + B^2 \eta \right\}
\end{aligned} \tag{2.7.18}$$

$$\begin{aligned}
im \{S_{i_g i_d}\} \Big|_{(0, x_0)} &= - \frac{16kT \omega C_{ox} WL}{(2 - \eta(L))^2 \eta(L)^2} \\
&\cdot \left\{ \frac{1}{4} A \eta^4 + \frac{1}{3} (B - 2A) \eta^3 + \frac{1}{2} (A - 2B) \eta^2 + B \eta \right\}
\end{aligned} \tag{2.7.19}$$

$$S_{i_d i_d} \Big|_{(0, x_0)} = \frac{4kT}{3} \cdot g_{d0} \cdot \frac{[1 - (1 - \eta)^3]}{\left[1 - \frac{1}{2} \eta(L)\right] \cdot \eta(L)}. \tag{2.7.20}$$

We simplify above three equations with: $g_{d0} = C_{ox} \left(\frac{W}{L}\right) \mu (V_{GS} - V_{TH})$. g_{d0} is the drain-to-source conductance when $V_{DS} = 0$. $\eta(x_0)$ is replaced by a simple η . Also,

$$\begin{aligned}
A &= \frac{2LI_0}{C_{ox} W \mu (V_{GS} - V_{TH})^2} \\
B &= \eta^2(L) - \frac{1}{3} \eta^3(L) - \frac{2LI_0 \eta(L)}{C_{ox} W \mu (V_{GS} - V_{TH})^2}.
\end{aligned} \tag{2.7.21}$$

Total noise contribution due to resistance between $(0, x_0)$ at saturation region

At saturation, the intrinsic FET satisfies $V_{DS} = V_{GS} - V_{TH}$, so:

$$\begin{aligned}
\eta(L) &= 1 \\
A &= 1 \\
B &= -\frac{1}{3}.
\end{aligned} \tag{2.7.22}$$

Equations (2.7.18), (2.7.19), and (2.7.20) will now become:

$$\begin{bmatrix} S_{i_g i_g} \\ im \{S_{i_g i_d}\} \\ S_{i_d i_d} \end{bmatrix} \Big|_{(0, x_0)} = \begin{bmatrix} \frac{32kT \omega^2 C_{ox}^2 W^2 L^2}{g_{d0}} \cdot \left\{ \frac{1}{5} \eta^5 - \frac{2}{3} \eta^4 + \frac{22}{27} \eta^3 - \frac{4}{9} \eta^2 + \frac{1}{9} \eta \right\} \\ -16kT \omega C_{ox} WL \cdot \left\{ \frac{1}{4} \eta^4 - \frac{7}{9} \eta^3 + \frac{5}{6} \eta^2 - \frac{1}{3} \eta \right\} \\ \frac{8kT g_{d0}}{3} \cdot [1 - (1 - \eta)^3] \end{bmatrix}. \tag{2.7.23}$$

Here, we express the aggregate noise contribution from the distributed resistors between $(0, x_0)$ inside the channel, when the FET is at saturation. η is a function of x_0 , and it is defined as $\eta(x_0) \equiv \frac{V_0(x_0)}{V_{GS} - V_{TH}}$. $V_0(x_0)$ is the voltage of the channel at location x_0 . Equation (2.7.23) is the basis of the Van der Ziel's noise trajectory in Figure 2.9, by plotting $\left[S_{i_g i_g}(\eta) \quad \text{im} \{ S_{i_g i_d}(\eta) \} \quad S_{i_d i_d}(\eta) \right]^T$ over $\eta = 0$ to 1.

Appendix 2.2: Pospieszalski's Noise Model

Pospieszalski assumes the noise in an FET is generated from the gate resistance at temperature T_g and the drain-to-source resistance at T_d . Since these two noise sources have different physical origins, they are uncorrelated. When $r \ll \frac{1}{\omega C_{gs}}$, it can be easily verified that the noise sources v_1 and i_2 have power spectral densities equal to $S_{v_1 v_1} = \frac{4}{5} \frac{kT\delta}{g_{d0}}$ and $S_{i_2 i_2} = 4kTg_{d0} \cdot \left(\gamma - \frac{\delta}{5} \right)$ to be able to fit $S_{i_g i_g}$ and $S_{i_d i_d}$ in Figure 2.8. Hence, noise source v_1 contributes the vector $\left[\frac{4}{5} \frac{kT\delta\omega^2 C_{gs}^2}{g_{d0}}, \frac{4}{5} \cdot kT\delta\omega C_{gs}, \frac{4}{5} \cdot kT\delta g_{d0} \right]^T$ and noise source i_2 contributes the other vector $\left[0, 0, 4kTg_{d0} \cdot \left(\gamma - \frac{\delta}{5} \right) \right]^T$ in the $\left[S_{i_g i_g}, \text{im} \{ S_{i_g i_d} \}, S_{i_d i_d} \right]^T$ vector space. The aggregate noise behavior of v_1 and i_2 is their vector sum.

Appendix 2.3: BSIM4 Noise Model

BSIM4 noise model fits the noise in an FET by a source noise voltage v_{bsim} and a drain noise current i_{bsim} , and they are assumed to be uncorrelated for ease of implementation. To fit the $S_{i_g i_g}$ and $S_{i_d i_d}$ in Figure 2.8, it can be easily verified that the power spectral density of $S_{v_{bsim} v_{bsim}} = \frac{4}{5} \frac{kT\delta}{g_{d0}}$ and $S_{i_{bsim} i_{bsim}} = 4kTg_{d0} \cdot \left(\gamma - \frac{\delta}{5}\right)$. Hence, noise source v_{bsim} contributes the vector $\left[\frac{4}{5} \frac{kT\delta\omega^2 C_{GS}^2}{g_{d0}}, \frac{4}{5} \cdot kT\delta\omega C_{gs}, \frac{4}{5} \cdot kT\delta g_{d0}\right]^T$ and noise source i_{bsim} contributes the other vector $\left[0, 0, 4kTg_{d0} \cdot \left(\gamma - \frac{\delta}{5}\right)\right]^T$ in the $\left[S_{i_g i_g}, \text{im}\{S_{i_g i_d}\}, S_{i_d i_d}\right]^T$ vector space. The aggregate noise behavior of v_{bsim} and i_{bsim} is their vector sum.

Chapter 3: A Compact Low Noise Weighted Distributed Amplifier

Signal amplification is the first and necessary step for the wireless receiver to recover communication signal from path loss (Figure 3.1). Since all physical amplifiers generate thermal noise, the amplification process also degrades the quality of signals. In order to compare the noise performance of different amplifiers, noise figure (NF) is defined as output signal-to-noise ratio (SNR) divided by input SNR ($NF = \frac{Output\ SNR}{Input\ SNR}$) for the purpose.

Using this definition, we can find that the overall NF of cascading system will be:

$$NF_{system} = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \dots \quad (3.1.1)$$

NF_i is the noise figure of the i -th stage, and G_i is the gain of the i -th stage. Since the gains of most blocks in communication system are normally greater than one, first-stage NF will dominate system noise performance.

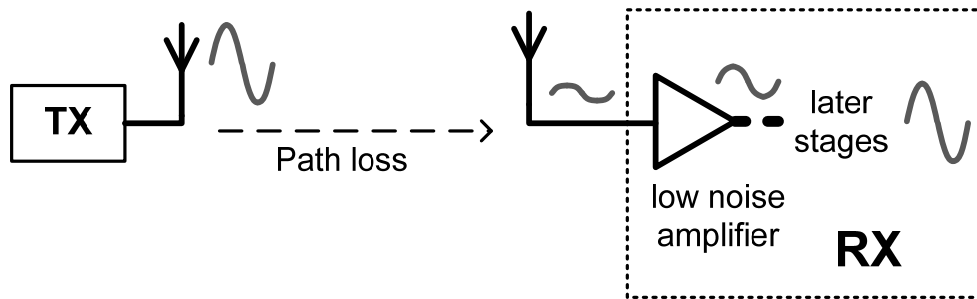


Figure 3.1: Signal strength in wireless communication

In addition to providing a low NF, the first amplifier also needs to provide a good input matching for the frontend antenna. This input matching requirement, however, poses a

design trade-off between the achievable bandwidth and the quality of the matching. We will first discuss this trade-offs for low noise amplifier (LNA) broadband matching in Section 3.1. Once input matching is achieved, to realize low noise operation will pose another challenge on power consumption for CMOS process, and this will be discussed in Section 3.2. In Section 3.3, we review the basic concept of distributed amplification (DA) which is capable of breaking the noise-bandwidth-power trade-offs. Different stages in a conventional DA contribute different noise to the amplifier's output. Using different weights for a DA will improve the noise performance of a conventional DA under the same power consumption. In Section 3.4, we discuss the noise process inside a weighted distributed amplifier (WDA), and a power-constraint noise optimization is carried in Section 3.5 to find the best weights. The use of many inductors to implement the LC-ladders in a conventional DA often result in a large layout area, and in Section 3.6, we introduce the coupling alternating LC-ladder to reduce its layout area. Schematics and the layouts of a WDA test chip will be discussed in Section 3.7. Experimental results of the test chip will be discussed in Section 3.8. Section 3.9 summarizes the highlights of this chapter.

3.1. Input Matching versus Bandwidth

An input matching network transforms particular impedance to the matched impedance over the design bandwidth. Fano [15] derives a criterion to determine the achievable matching of physically realizable networks. In Figure 3.2, we use a commonly used special case to explain his idea. Assuming that a LNA has an equivalent circuit equal to a

impedance R_{in} and C_{in} , and we want to design a broadband matching network to match the parallel RC to a constant impedance, the achievable matching of any physically realizable network to match the parallel RC will have to satisfy:

$$\rho(\omega) \equiv \frac{P_{ref}(\omega)}{P_{in}(\omega)} \quad (3.1.2)$$

$$\int_0^\infty \ln\left(\frac{1}{\rho(\omega)}\right) \cdot d\omega \leq \left(\frac{\pi}{C_{in} \cdot R_{in}}\right).$$

$\rho(\omega)$ is the reflection coefficient, $P_{ref}(\omega)$ is the reflected power, and $P_{in}(\omega)$ is the input power. If we have such a parallel RC to match, and we match the LNA RC input to a fixed ρ_0 over a bandwidth (BW), applying this example to Equation (3.1.2), and we will get:

$$C_{in} \leq \frac{\pi}{R_{in} \cdot \ln\left(\frac{1}{\rho_0}\right) \cdot BW}. \quad (3.1.3)$$

This means that in order to match the LNA over the design BW with constant reflection coefficient ρ_0 , the equivalent parasitic capacitance C_{in} of the LNA input needs to be smaller than Equation (3.1.3). In most design cases, C_{in} is contributed from the active device and the metal interconnections. Since the input parasitic capacitance of an active device is proportional to its device sizing, Equation (3.1.3) also implies that we cannot use an arbitrary large device, or the LNA won't achieve the required BW. In addition, in BW the center frequency is a design parameter, so designing a high-frequency narrow bandwidth LNA can be easier than a low frequency broadband LNA. Furthermore, LNA is an I/O block, and electro-static discharge (ESD) protection is necessary. Applying ESD protection in a broadband LNA will consume this C_{in} budget.

The parallel RC example in Figure 3.2 is a special case that leads to Equation (3.1.2). The general case, though mathematically laborious, is discussed in Fano's thesis [15].

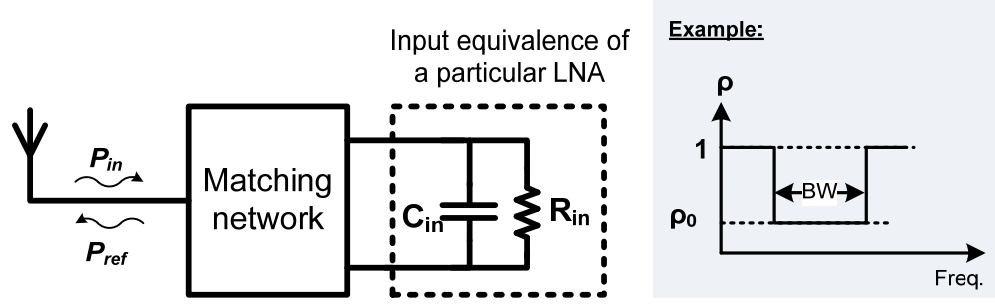


Figure 3.2: Trade-offs for input matching and input parasitic capacitance

3.2. Issues of Power-Constraint LNA Optimization in CMOS

In Section 3.1, we conclude that large active device cannot be used in the design of a broadband LNA. An LNA also needs to achieve low noise operation. The NF of any LNA is a function of both the effective transconductance G_m and the part contributed from other noise sources:

$$NF = \frac{Noise|_{R_0} + Noise|_{G_m} + other\ noise}{Noise|_{R_0}}. \quad (3.2.1)$$

$Noise|_{R_0}$ is the output noise due to the input termination resistor R_0 , $Noise|_{G_m}$ is the output noise due to the active device, and *other noise* is the total output noise contribution from all other parts. $Noise|_{R_0} \approx C_1 \cdot kTR_0G_m^2$, because the power gain of an amplifier is proportional to G_m^2 . $Noise|_{G_m} \approx C_2 \cdot G_m$, because the noise generated from the active device is roughly proportional to its transconductance. The *other noise* term is a weak function of G_m . Both C_1 and C_2 is a constant design parameter for a given design. We can rewrite Equation (3.2.1) as:

$$\begin{aligned}
NF &\approx \frac{C_1 \cdot kTR_0 G_m^2 + C_2 \cdot G_m + \text{other noise}}{C_1 \cdot kTR_0 G_m^2} \\
&\approx \frac{C_1 kTR_0 + C_2 \frac{1}{G_m} + \frac{\text{other noise}}{G_m^2}}{C_1 kTR_0}.
\end{aligned} \tag{3.2.2}$$

We can reduce the noise figure of a noise figure by increasing G_m . Since parasitic capacitance from the active device in a broadband LNA has an upper limit, the only way to improve its NF is to bias the active device at high $f_T = \frac{g_m}{C_{gs}}$ region. Biasing a CMOS transistor at high f_T corresponds to a low $\frac{g_m}{I_{bias}}$, as shown in Figure 3.3. In other words, in order to get a large g_m , we need to consume a much larger current compared to the intrinsic bipolar junction transistor (BJT). In addition, for a given g_m , BJT shows less current noise at its drain node compared to a short-channel CMOS transistor. This trade-off prevents CMOS broadband LNA from low-power application compared with its BJT counterpart.

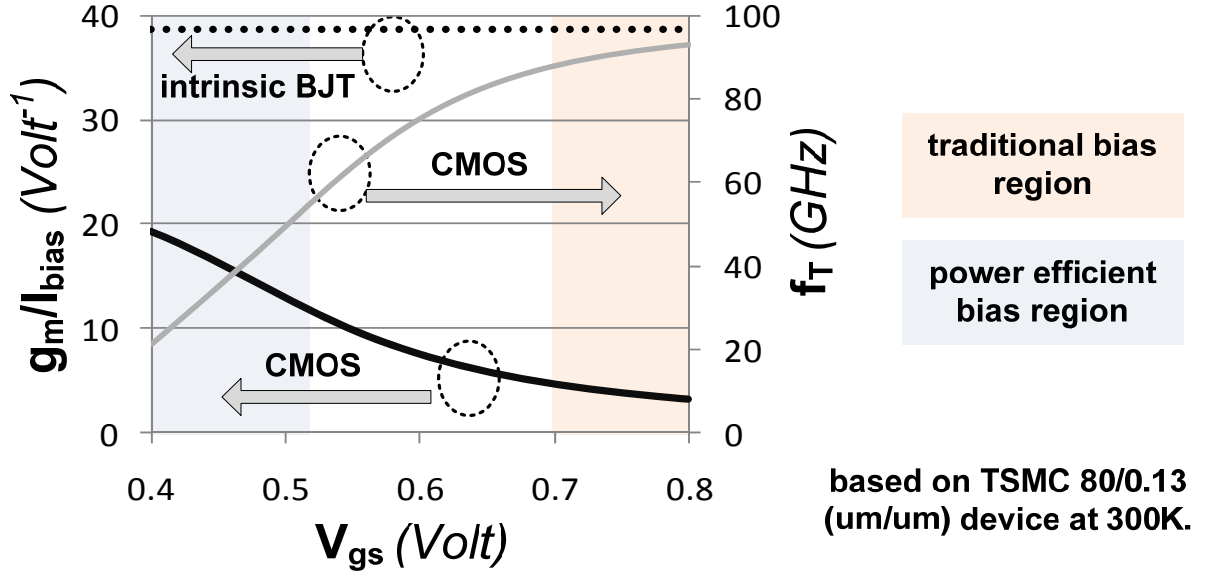


Figure 3.3: Transconductance current efficiency and f_T versus bias voltage of a CMOS transistor and an intrinsic BJT

3.3. Low-Noise Distributed Amplifier

The low noise and power consumption trade-off mentioned in Section 3.2 can be broken by using distributed amplification. A DA connects several parallel gain stages at their inputs and outputs by inter-stage inductors as shown in Figure 3.4. Parasitic input and output capacitance of the gain stages will be absorbed into the effective input and output LC-ladders. If we terminated one end of the artificial LC-ladder with the resistor equal to its intrinsic impedance ($Z_0 = \sqrt{\frac{L_{in}}{C_{in}}}$), and equal the sectional group delays ($\Delta T = \sqrt{L_{in}C_{in}}$) for input and output ladders, output current from each of the gain stages will be combined in phase up to the ladder bandwidth.

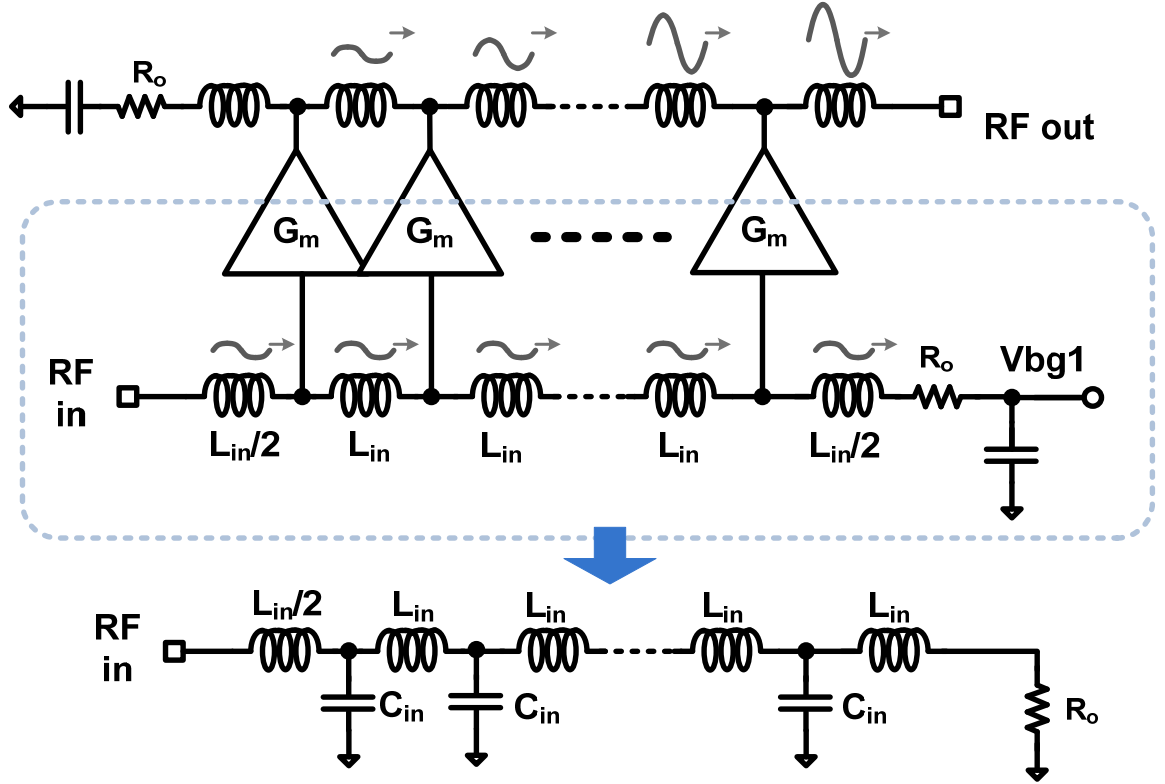


Figure 3.4: The concept of distributed amplification

The bandwidth of a DA is determined by the bandwidth of both of the LC-ladders, which have a cut-off frequency $f_c = \frac{1}{\pi\sqrt{L_{in}C_{in}}}$. The total parasitic capacitance budget that can be absorbed by the LC-ladders is now $N \cdot C_{in}$, and the total effective transconductance of the DA is now $N \cdot G_m$. Here, N is the number of stages. If the LC-ladders are lossless, there are no limitations on the number of stages. The loss in LC-ladders will attenuate the wave propagating along the ladders, and will reduce the benefits of distributed amplification for large N .

Since both the total parasitic capacitance and the effective transconductance are increased by N -times, this means that even if each stage has a low stage G_m , we can still

achieve a large total effective transconductance by increasing the number of stages. So, we can bias the transistors inside each gain stage at their low f_T region to reduce the power consumption of the overall DA. The DA will have a good noise figure due to the large effective total G_m . Low-noise DA based on power-constraint optimization has been studied by Hedari [16].

There are several issues about previous research on DA: first of all, the noise of each stage contributes different noise power to the output. If we can change the weights of each gain stage, as opposed to the uniform weights in the conventional DA, we have an additional design dimension to improve its overall noise figure within the design bandwidth. Secondly, the noise studies of a conventional DA are based on classic transistor model which has correlated gate and drain noise. As mentioned in Chapter 2, this noise correlation causes noise analysis of large networks too complex to get design insights. We will use the new noise model to re-study its noise performance in the next section.

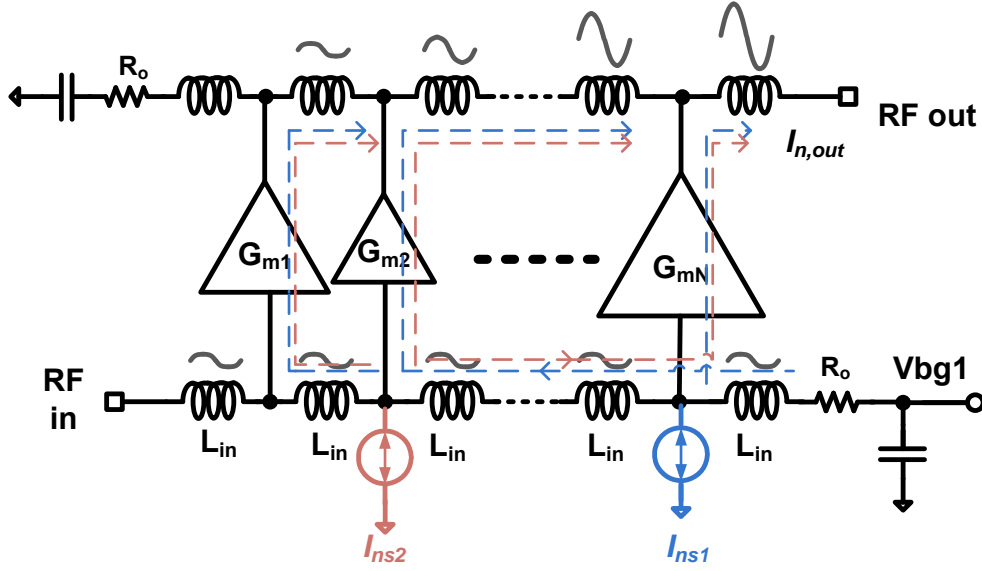


Figure 3.5: The concept of weighted distributed amplifier

3.4. Noise Process in the Weighted Distributed Amplifier (WDA)

A weighted distributed amplifier (WDA) differs from a conventional uniform DA by its non-uniform gain stages as: G_{m1} , G_{m2} , ..., G_{mN} (shown in Figure 3.5). Since each stage is weighted, thermal noise generation is different from each stage. In addition, the weighted stages form an effective finite-impulse-response noise filtering system for different noise sources. For example, output noise due to noise sources I_{n1} and I_{n2} in Figure 3.5 is:

$$\begin{aligned} \overline{|I_{n,out}|^2} &= \left| \sum_{i=1}^N G_{mi} \cdot e^{-2\gamma(N-i)} \right|^2 \cdot \overline{|I_{ns1}|^2} + \\ &\left| \sum_{i=1}^2 G_{mi} \cdot e^{-2\gamma(2-i)-\alpha(N-2)} + \sum_{i=3}^N G_{mi} e^{-\alpha(N-2)} \right|^2 \cdot \overline{|I_{ns2}|^2}. \end{aligned} \quad (3.4.1)$$

$\gamma = \alpha + j\omega T_g$ is the propagation constant of a LC-section of the ladder, where α is the sectional attenuation constant, and T_g is the sectional group delay. It is clear from Equation

(3.4.1) that the transfer functions of I_{ns1} to $I_{n,out}$ and from I_{ns2} to $I_{n,out}$ have different finite-impulse responses. The finite-impulse responses are function of both the stage weights and location of the noise sources. The whole WDA can be viewed as a complex finite-impulse-response system. We will study the noise transfer functions of different noise sources inside the WDA.

3.4.1. Noise from Common-Source Transistors

Conventional noise calculation utilizes a noise model with two correlated sources. This results in an analytic formula that contains a long algebraic multiplication term due to this noise correlation. We will use the modified FET noise model introduced in Chapter 2 for noise analysis and calculation.

In Figure 3.6, the drain noise of the i -th common-source transistor contributes to the output through two parts. The major part comes from the direct amplification by its cascode transistor, and:

$$I_{out_dx,i_major} = \frac{1}{2}\eta \cdot e^{-\gamma(N-i)} \cdot I_{dx,i}. \quad (3.4.2)$$

η is the current efficiency from the small signal transconductance (g_m) to the drain output of the cascode transistor. $\eta \approx \frac{g'_{m,i}}{g_{d,i} + g'_{m,i}}$, $g'_{m,i}$ is the transconductance of the cascode transistor, and η is very close to unity. $\gamma = \alpha + j\omega T_g$, where α is the sectional attenuation constant and T_g is the sectional group delay. The factor of $\frac{1}{2}$ in Equation (3.4.2) is due to half of the output current from the cascode transistor being split into a backward propagating wave, and not contributing to LNA output noise.

A small portion of this noise leaks to the input LC-ladder through C_{gd} , and this leakage noise propagates both forward and backward in the input LC-ladder, and is amplified by all stages other than $G_{m,i}$. Since C_{gd} is usually very small, the leakage noise is very small. However, it is also amplified by lots of stages, so its contribution can be important.

$$I_{out_dx,i_minor} \approx \frac{1}{4} Z_0 \eta \cdot (j\omega C_{gd,i}) \cdot \frac{I_{dx,i}}{g_{m,i}} \cdot e^{-\gamma(N-i)} \cdot \left\{ \sum_{k=1}^{i-1} g_{m,k} \cdot e^{-2\gamma(i-k)} + \sum_{k=i+1}^N g_{m,k} \right\} \quad (3.4.3)$$

Adding Equations (3.4.2) and (3.4.3) we get:

$$I_{out_dx,i} = \frac{1}{2} \eta e^{-\gamma(N-i)} \left\{ 1 + \frac{j\omega C_{gd,i} Z_0}{2g_{m,i}} \left[\sum_{k=1}^{i-1} g_{m,k} \cdot e^{-2\gamma(i-k)} + \sum_{k=i+1}^N g_{m,k} \right] \right\} \cdot I_{dx,i}. \quad (3.4.4)$$

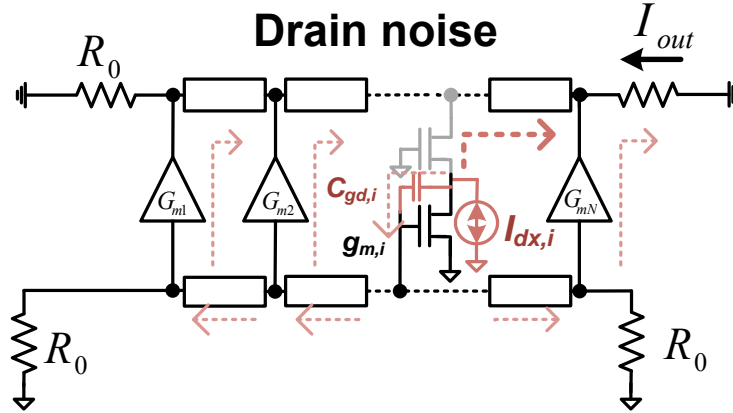


Figure 3.6: Noise from the drain noise of the i -th common-source transistor

Gate noise from the i -th common-source transistor contributes a forward and backward wave in the input LC-ladder, as shown in Figure 3.7.

$$I_{out_gx,i} = \frac{1}{2} \eta e^{-\gamma(N-i)} \left[\sum_{k=i}^N g_{m,k} + \sum_{k=1}^{i-1} g_{m,k} \cdot e^{-2\gamma(i-k)} \right] \cdot I_{gx,i}. \quad (3.4.5)$$

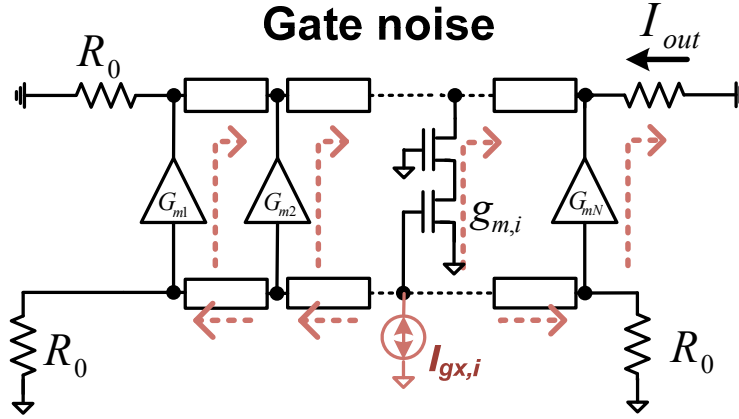


Figure 3.7: Noise from the gate noise of the i -th common-source transistor

Similar to the drain noise, the source noise has a part that is directly amplified by the cascode amplifier, and another part amplified by other stages due to noise leakage to the input LC-ladder.

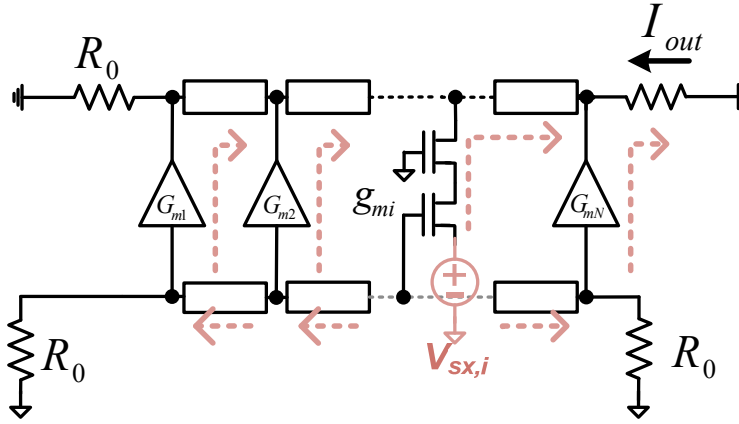


Figure 3.8: Noise from the source noise of the i -th common-source transistor

$$I_{out_{sx,i}} = \frac{1}{2} \eta e^{-\gamma(N-i)} \left\{ g_{m,i} - \frac{1}{2} j \omega C_{gs,i} Z_0 \left[\sum_{k=1}^{i-1} g_{m,k} e^{-2\gamma(i-k)} + \sum_{k=i}^N g_{m,k} \right] \right\} \cdot V_{sx,i} \quad (3.4.6)$$

The total noise from the i -th CS transistor can be derived by squaring the absolute value of Equations (3.4.4), (3.4.5), and (3.4.6); we get:

$$\overline{|I_{out_{cs,i}}|^2} = \overline{|I_{out_{dx,i}}|^2} + \overline{|I_{out_{gx,i}}|^2} + \overline{|I_{out_{sx,i}}|^2}. \quad (3.4.7)$$

And the total noise due to all the common source transistors will be:

$$\overline{|I_{out}|^2}_{cs} = \sum_{i=1}^N \overline{|I_{out_{cs,i}}|^2}. \quad (3.4.8)$$

3.4.2. Noise from Cascode Transistors

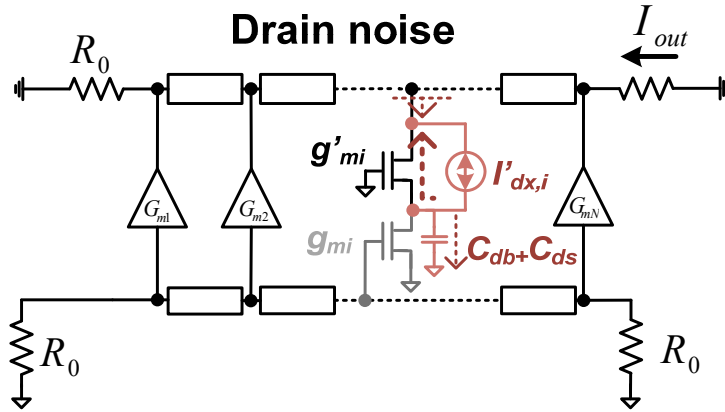


Figure 3.9: Noise from the drain noise of the i -th cascode transistor

The cascode transistor can be analyzed in a similar way as the CS transistor. In Figure 3.9, the noise transfer function from the drain noise of the i -th cascode transistor is illustrated. The noise from the cascode transistor is not important at lower frequency due to the high impedance at the drain node of the CS transistor. For higher frequency, the parasitic drain-to-bulk capacitor (C_{db}) and the drain-to-source capacitor (C_{ds}) shunt with the CS transistor's output conductance (g_d), so the noise from the cascode transistor increases with frequency. This drain noise's contribution to the output will be:

$$I_{out_{dx,i}} \approx \frac{1}{2} \cdot \left[\frac{j\omega(C_{db,i}+C_{ds,i}+C_{gd,i})}{g'_{m,i}+j\omega(C_{db,i}+C_{ds,i}+C_{gd,i})} \right] \cdot e^{-\gamma(N-i)} \cdot I'_{dx,i}. \quad (3.4.9)$$

The gate noise's contribution to the output is similar to the drain noise, and the only difference is that the other side of the gate noise current source is shorted with an RF ground instead of a constant impedance. Its noise contribution to output is:

$$I_{out_{gx,i}} \approx \frac{1}{2} \cdot \left[\frac{j\omega(C_{db,i}+C_{ds,i}+C_{gd,i})}{g'_{m,i}+j\omega(C_{db,i}+C_{ds,i}+C_{gd,i})} \right] \cdot e^{-\gamma(N-i)} \cdot I'_{gx,i}. \quad (3.4.10)$$

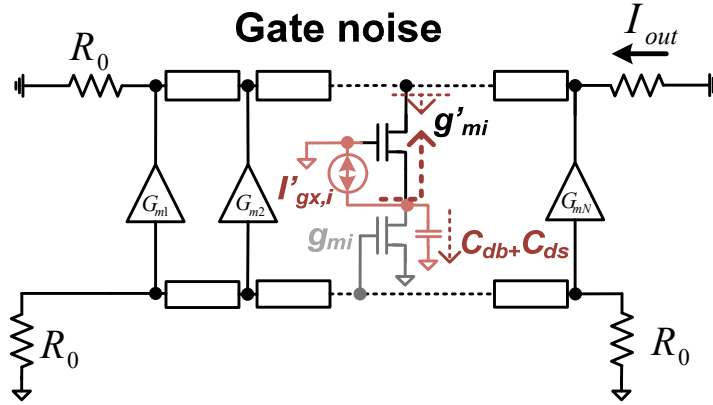


Figure 3.10: Noise from the gate noise of the i -th cascode transistor

The source noise voltage is in series with the source impedance ($\frac{1}{g'_{m,i}}$) and the parasitic capacitance at the drain of the CS transistor, so its noise contribution to the output is:

$$I_{out_{sx,i}} \approx \frac{1}{2} \cdot \left[\frac{j\omega \cdot g'_{m,i} \cdot (C_{db}+C_{ds}+C_{gd})}{g'_{m,i}+j\omega(C_{db}+C_{ds}+C_{gd})} \right] \cdot e^{-\gamma(N-i)} \cdot V'_{sx,i}. \quad (3.4.11)$$

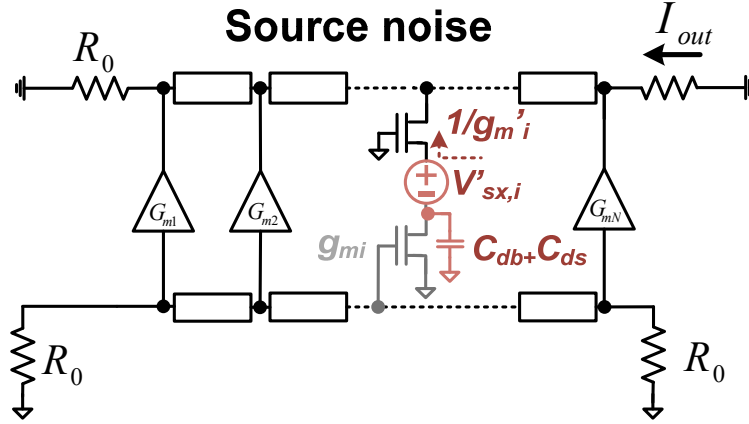


Figure 3.11: Noise from the source noise of the i -th cascode transistor

Similarly, total noise from the i -th cascode transistor can be derived by squaring the absolute value of Equations (3.4.9), (3.4.10), and (3.4.11); we get:

$$\overline{|I_{out_{cascode,i}}|^2} = \overline{|I_{out_{dx',i}}|^2} + \overline{|I_{out_{gx',i}}|^2} + \overline{|I_{out_{sx',i}}|^2}. \quad (3.4.12)$$

And the total noise due to all the common source transistors will be:

$$\overline{|I_{out}|^2}_{cascode} = \sum_{i=1}^N \overline{|I_{out_{cascode,i}}|^2}. \quad (3.4.13)$$

3.4.3. Noise from Termination Resistors

Both the input and the output LC-ladders have termination resistors, which contribute noise to the output. The output noise current due to the termination resistor of the input LC-ladder is:

$$\overline{|I_{out}|^2}_{in,term} = kTR_{in,term} \cdot \left| \left\{ \sum_{k=1}^N \eta g_{m,k} \cdot e^{-2\gamma(N-k)} \right\} \right|^2. \quad (3.4.14)$$

$R_{in,term}$ is the input LC-ladder's termination resistance. The output noise current due to the termination resistor of the output LC-ladder is:

$$\overline{|I_{out}|^2}|_{out,term} = \frac{kT}{R_{out,term}}. \quad (3.4.15)$$

$R_{out,term}$ is the output LC-ladder's termination resistance.

3.4.4. Noise from Passive Network Loss

All physical LC-ladders are lossy, and they generate thermal noise to the network. There are several ways to model this loss and the associated noise. One simple way is to attribute the loss to inductors, and model the inductors using a frequency-dependent model, which consists of an ideal inductor in series with a frequency-dependent loss resistor $r_{loss}(\omega)$. The equivalent circuit of the LC-ladder with a lossy inductor between the $(i-1)$ -th and the i -th stage is shown in Figure 3.12.

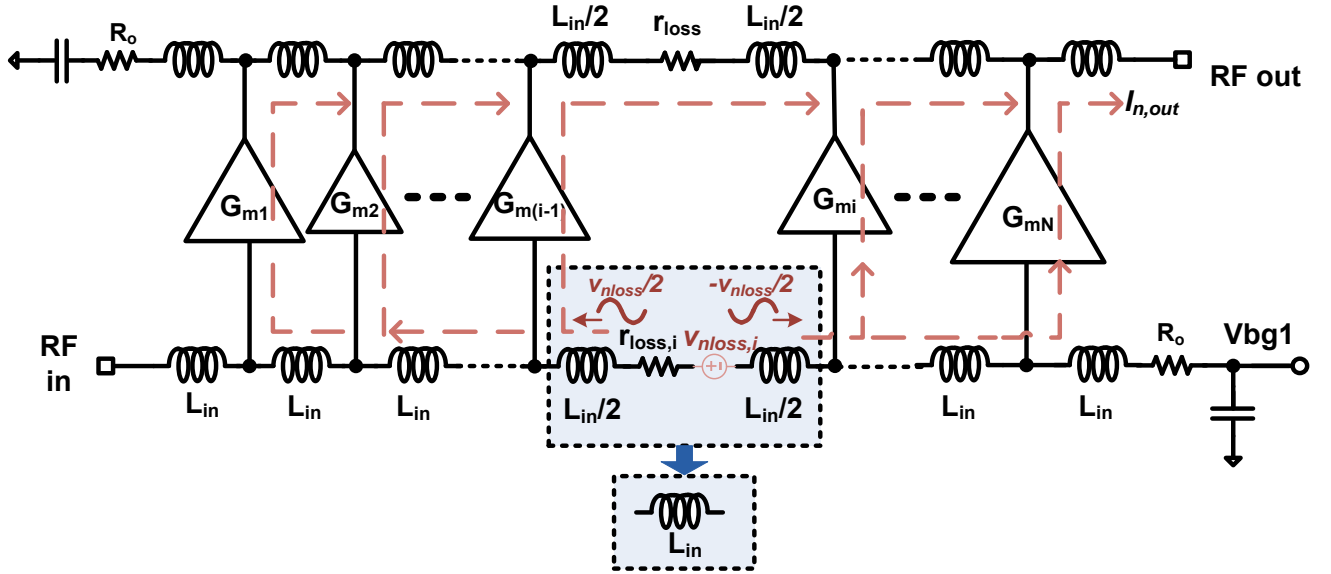


Figure 3.12: Equivalent circuit of the LC-ladder with a lossy inductor between the $(i-1)$ -th and i -th stage

So r_{loss} generates a noise voltage that gives forth a forward voltage wave and a backward voltage wave which has opposite magnitude. The total output noise due to the loss resistor between the $(i-1)$ -th and the i -th stage will be:

$$I_{out,rloss,i} \approx \frac{1}{2} e^{-(N-i+1)\gamma} \cdot \eta \cdot \left[\sum_{k=1}^{i-1} g_{m,k} \cdot e^{-2\gamma(i-1-k)-\gamma} - \right. \\ \left. k=ik=Ngm,k \cdot v_{nloss,i} \right] \quad (3.4.16)$$

Based on Equation (3.4.16), the thermal noise from this loss resistance is important for the first several inductors, because most of its noise has been amplified in phase. The noise from the later stages is FIR filtered, and their significance is reduced. In addition, the loss resistor increases with frequency due to the skin effects of the interconnection metals. Post-layout 3-D EM simulation is necessary to estimate the loss of the LC-ladders. Since $r_{loss}(\omega)$ is consequence of the inductor layout, different inductor layouts result in different inductance values. Under the constraints of constant ladder impedance (50Ω), the change of inductor layout affects both the sectional group delay and $r_{loss}(\omega)$. If we use this relationship and let sectional group delay of the LC-ladder be a design parameter, $r_{loss}(\omega)$ will be a function of group delay during the optimization process. The total noise from the input loss resistors will be:

$$\overline{|I_{out}|^2}_{r_{loss}} = \sum_{i=2}^N \overline{|I_{out,rloss,i}|^2}. \quad (3.4.17)$$

The output LC-ladder also contributes thermal noise, and the total output noise due to the loss resistor between the $(i-1)$ -th and the i -th stage will be:

$$I_{out,rloss,o,i} \approx \frac{1}{2Z_0} \cdot e^{-(N-i+1)\gamma} \cdot v_{rloss,i}. \quad (3.4.18)$$

So, the total noise due to the output LC-ladder is:

$$\overline{|I_{out}|^2}_{r_{loss,o}} = \sum_{i=2}^N \overline{|I_{out,r_{loss,o,t}}|^2}. \quad (3.4.19)$$

3.4.5. Voltage Peaking Effect in LC-Ladder

When we drive a uniform LC-ladder with a broadband power source, though the input power is constant over the frequency range, the amplitude of the generated voltage wave propagating along the LC-ladders varies with frequency, and its amplitude versus frequency is:

$$\frac{V_{g1}(s)}{V_{g1}(0)} = \frac{4}{4+2s\Delta t+s^2\Delta t^2}. \quad (3.4.20)$$

$V_{g1}(s)$ is the amplitude of the voltage wave at frequency ω , where $s = j\omega$. $\Delta t = \sqrt{LC}$, and this response has a unity quality factor. Equation (3.4.20) has a slight voltage peaking effect at higher frequency, and we can use this favorable effect to compensate the increasing noise due to the cascode transistors and the loss resistors inside the LC-ladders.

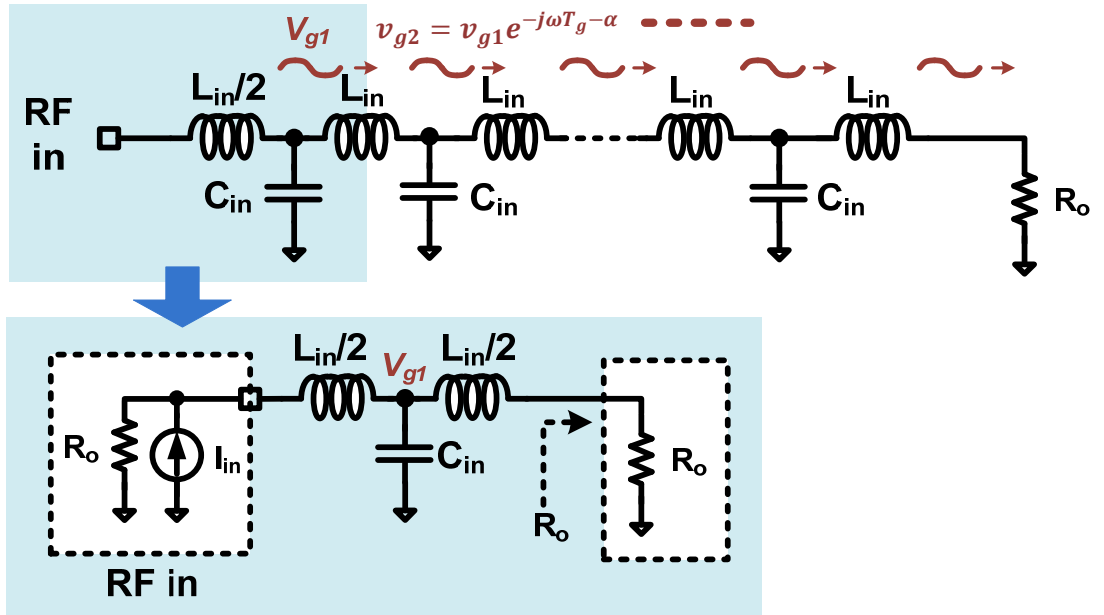


Figure 3.13: Driving the LC-ladders with a broadband power source, and its equivalent circuit

There is a similar effect when we drive the LC-ladder from an internal node with a broadband current source, as shown in Figure 3.14. This is the situation when we connect the output of a gain stage to the LC-ladder. The magnitude of the voltage wave has a frequency response of:

$$\frac{V_{g1}(s)}{V_{g1}(0)} = \frac{4}{4 + 2s\Delta t + s^2\Delta t^2}. \quad (3.4.21)$$

This equation is the same as Equation (3.4.20). We can use both input and output voltage peaking to improve our design.

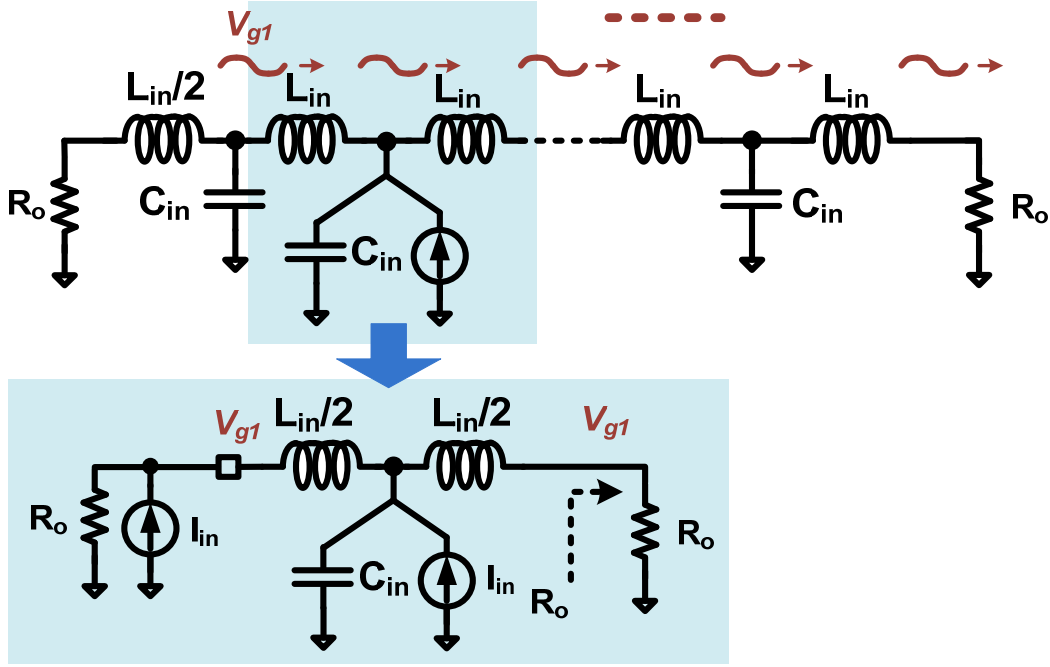


Figure 3.14: Driving the output LC-ladder from an internal load, and its equivalent circuit

3.4.6. Frequency-Dependent Group Delay

A regular LC-ladder has a frequency-dependent sectional group delay. At low frequency, the sectional phase delay is:

$$\phi = 2\pi f \Delta t_0 = \frac{2\pi f}{\pi f_c} = \frac{f}{f_c}, \text{ for } f \ll f_c. \quad (3.4.22)$$

When the signal frequency approaches the cut-off frequency of the ladder, this phase shift will show strong frequency dependency:

$$\phi(f) = \text{imag} \left\{ \ln \left(1 - 2 \cdot \left(\frac{f}{f_c} \right)^2 + \frac{2f}{f_c} \sqrt{\left(\frac{f}{f_c} \right)^2 - 1} \right) \right\} > 2 \cdot \frac{f}{f_c}. \quad (3.4.23)$$

And the effective group delay is defined as: $\Delta t_{eff}(f) \equiv \frac{\phi(f)}{2\pi f}$, where f is the frequency of interest.

The noise process inside a WDA can be viewed as a FIR system for different noise sources, and the changes in group delays affect the noise impulse responses of each noise source. This frequency-dependent group delay needs to be considered in noise analysis and optimization.

3.4.7. Frequency-Dependent Impedance Change

A LC-ladder has frequency-dependent intrinsic impedance. This effect causes frequency-dependent signal reflection at the both ends of a LC-ladder. The reflected signals will further propagate along LC-ladders until they fade away. Due to these reflections, impulse response of a given noise source changes accordingly. To take the signal reflections into account, we first calculate the intrinsic impedance at frequency f :

$$Z_{int}(f) = R_0 \sqrt{1 - \frac{f^2}{f_c^2}}. \quad (3.4.24)$$

$Z_{int}(f)$ is the intrinsic impedance, and R_0 is the impedance at DC. The reflection coefficients at both ends of the LC-ladder will be:

$$\frac{b}{a} = \frac{R_{term} - Z_{int}(f)}{R_{term} + Z_{int}(f)}. \quad (3.4.25)$$

b is the normalized reflected wave from the termination, for an input wave a . The reflected wave can have a same or opposite magnitude of the input wave.

3.4.8. Noise Figure of WDA

After applying the second-order effects into Equations: (3.4.8), (3.4.13), (3.4.14), (3.4.15), (3.4.17), and (3.4.19), and also knowing that the output noise due to the source impedance is

$$\overline{|I_{out}|^2}|_{src,term} = kTR_{src,term} \cdot \left| \left\{ \sum_{k=1}^N \eta g_{m,k} \cdot e^{-\gamma N} \right\} \right|^2, \quad (3.4.26)$$

we get the total output noise current

$$\begin{aligned} \overline{|I_{out}|^2}|_{total} = & \overline{|I_{out}|^2}|_{src,term} + \overline{|I_{out}|^2}|_{CS} + \overline{|I_{out}|^2}|_{cascode} + \\ & \overline{|I_{out}|^2}|_{in,term} + \overline{|I_{out}|^2}|_{out,term} + \overline{|I_{out}|^2}|_{rloss} + \overline{|I_{out}|^2}|_{rloss,o}, \end{aligned} \quad (3.4.27)$$

and the noise figure of the WDA is thus

$$NF = \frac{\overline{|I_{out}|^2}|_{total}}{\overline{|I_{out}|^2}|_{src,term}}. \quad (3.4.28)$$

Equation (3.4.28) is a function of stage weights, ladder group delays, total bias current, and transistor gate bias voltage. We will use this noise analysis and calculation for the WDA optimization.

3.5. Power-Constraint Noise Optimization of WDA

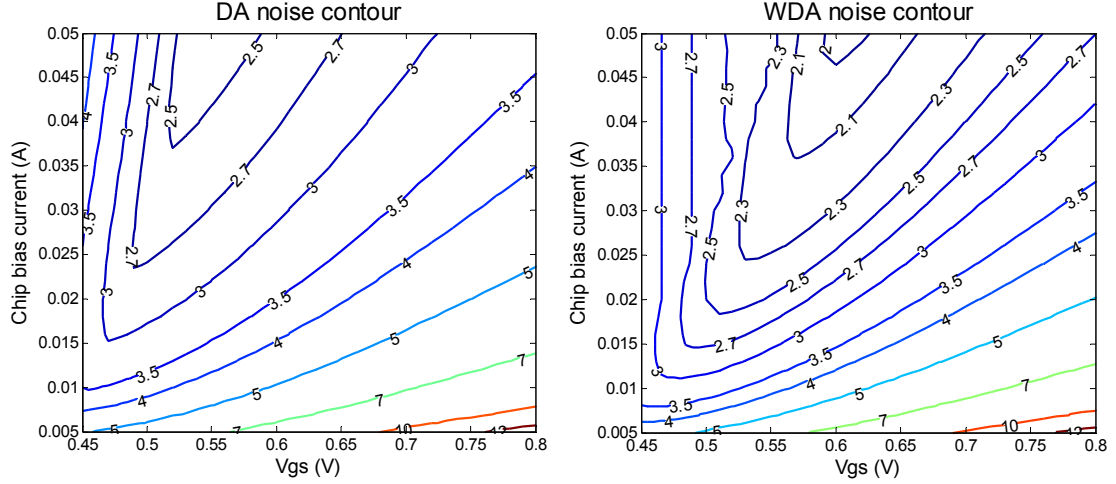


Figure 3.15: Power-constraint noise optimization contour comparisons between a DA and a WDA

Based on the discussions in Section 3.4, we can calculate the noise figure of a WDA as a function of a stage weights, ladder group delays, total bias current, and transistor gate bias voltage. We first constrain the intrinsic impedance of the LC-ladder to be $50\ \Omega$ ($Z_0 = \sqrt{\frac{L_{in}}{C_{in}}}$) for the ease of RF measurement. Secondly, we constrain the WDA bandwidth to cover at least $10.6\ GHz$ ($f_c = \frac{1}{\pi T_g} > 10.6\ GHz$); this suggests an upper limit for the choice of sectional group delay ($T_g < \frac{1}{\pi \cdot 10.6\ GHz}$). This also implies that the maximum parasitic capacitance from both the input and the output of a stage amplifier is upper bounded to ($C_{in} \leq \frac{1}{\pi Z_0 f_c}$). We can carry a power-constraint noise optimization based on these two constraints:

Optimization Algorithms:

1. For each total current consumption, and CS-transistor gate bias voltage:

a. Sweep $f_c > 10.6$ GHz.

i. Calculate sectional group delay $(T_g = \frac{1}{\pi f_c})$.

ii. Use the constraints of maximum stage parasitic capacitance

$$(C_{in} \leq \frac{1}{\pi Z_0 f_c}) \text{ and } \underline{\text{CS-transistor gate bias voltage}}$$

to calculate the maximum transistor width (W_{\max}) for each gain stage. Also, use

the (total current consumption, CS-transistor gate bias voltage) to

calculate total transistor width of all stages ($W_{\text{total}} = \sum_{i=1}^N W_i$).

Based on this, we can also calculate a corresponding maximum

transconductance of the CS-transistor ($G_{m,\max}$) of each gain stage,

and the total transconductance of the CS-transistors ($G_{m,\text{total}}$).

iii. Use fminimax algorithm in matlab [61] to optimize stage

transconductance $G_{m,i}$ for goal (NF), based on Equation (3.4.28)

with constraints $G_{m,i} \leq G_{m,\max}$ and $\sum_{all i} G_{m,i} = G_{m,\text{total}}$.

iv. Save the optimization result $G_{m,i}|_{\text{optimized},f_c}$ and optimized goal

$NF|_{\text{optimized},f_c}$.

- b. Save the best optimization result $G_{m,i}|_{optimized}$ and goal $NF|_{optimized}$ among all f_c in consideration.
2. Save the goal $NF|_{optimized}$ for the given (total current consumption, CS-transistor gate bias voltage) pair.

We choose our design goal to be achieving *minimum worst-case noise figure between 3.1–10.6 GHz*. The optimization process is based on the 130 nm CMOS device model provided by TSMC. In Figure 3.15, power-constraint noise optimization comparison results between a conventional DA and an optimized WDA are plotted. For a given total IC current consumption and a given gate bias voltage, the optimized WDA achieves a better in-band worst-case noise figure compared to a conventional DA. This demonstrates the added design flexibility of a WDA improves the noise performance of a conventional DA.

3.6. Magnetic Couplings in LC-Ladder

The conventional LC-ladders have large layout area due to the use of many inductors. To avoid couplings between adjacent inductors, it is a common design practice to place inductors far apart from each other, and this worsens the situation. If we apply magnetic couplings between adjacent inductors in a LC-ladder as shown in Figure 3.16, and formulate the circuit *KCL* and *KVL* equations, we will get:

$$\begin{aligned}
 v(x) - v(x + \Delta x) &= i(x) \cdot sL + i(x - \Delta x) \cdot sM + i(x + \Delta x) \cdot sM \\
 v(x) \cdot sC + i(x) &= i(x - \Delta x).
 \end{aligned}
 \tag{3.5.1}$$

If we simplify Equation (3.5.1) using linear approximation we will get:

$$\begin{aligned}
 -\frac{dV}{dx} &= i(x) \cdot s \left\{ \frac{L}{\Delta x} + \frac{2M}{\Delta x} \right\} \\
 -\frac{dI}{dx} &= V(x) \cdot s \left\{ \frac{C}{\Delta x} \right\}.
 \end{aligned}
 \tag{3.5.2}$$

Solving Equation (3.5.2) we will get the intrinsic impedance and the sectional group delay of the adjacently coupling LC-ladder:

$$Z_0 = \sqrt{\frac{L + 2M}{C}} \tag{3.5.3}$$

$$T_{g,section} = \sqrt{(L + 2M)C}.$$

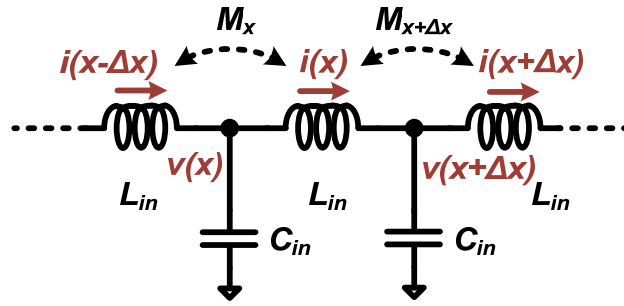


Figure 3.16: Adjacent couplings in a LC-ladder

The solutions of Equation (3.5.2) reveal interesting design insights. First of all, though signals at different inductors have different signal phases due to group delay, the phase lead of the inductor on the left cancels the phase lag of the inductor on the right to the first order. As a result, the overall coupling LC-ladder works in a similar way as a noncoupling LC-ladder with a change on the intrinsic impedance (Z_0) and the group delay ($T_{g,section}$)

as written in Equation (3.5.3). A constructive coupling ($M > 1$) increases both the effective intrinsic impedance and the group delay and vice versa.

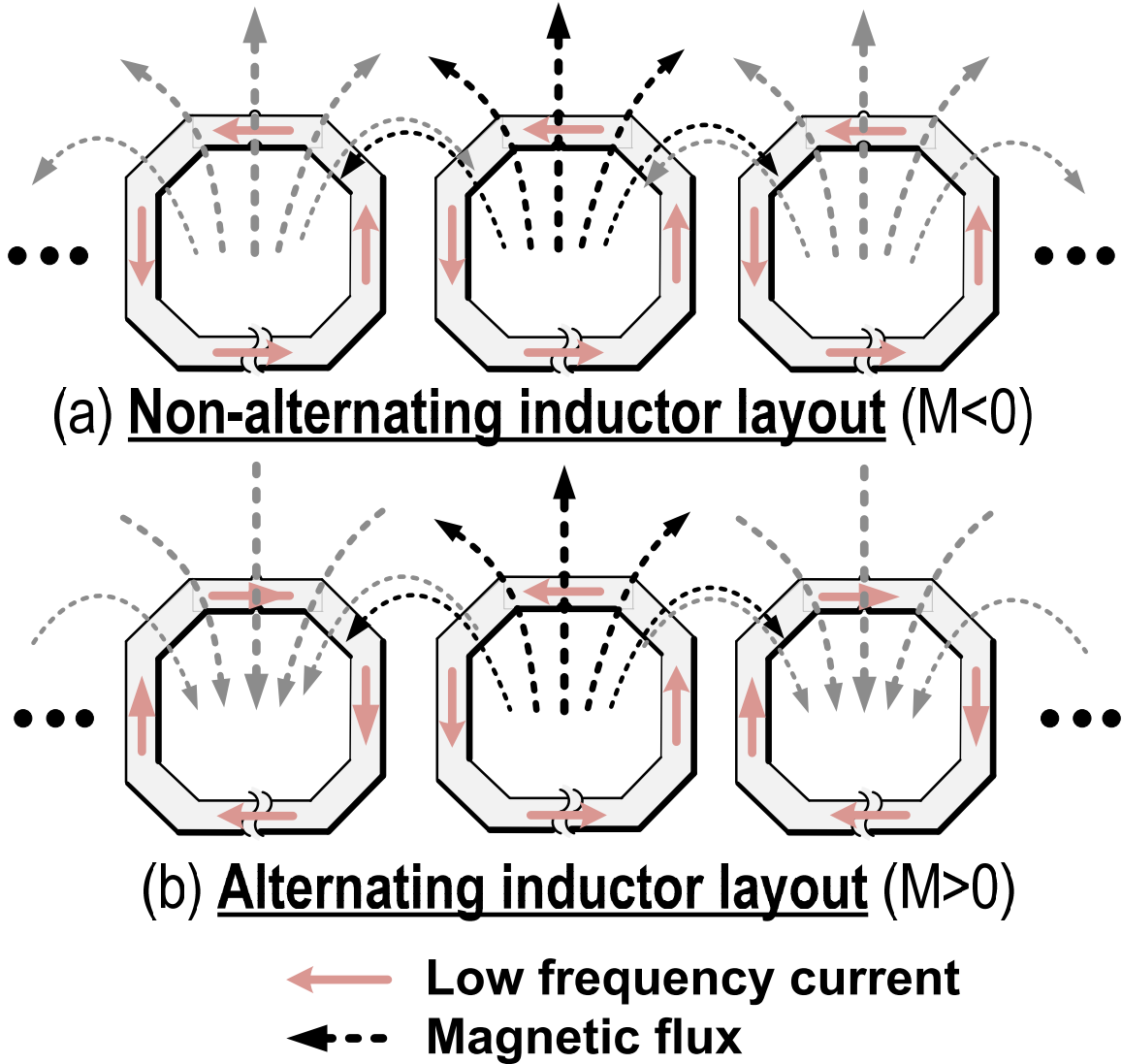


Figure 3.17: Inductor layouts in two different LC-ladders: (a) Non-alternating inductor layout, and (b) Alternating inductor layout

The direction of magnetic couplings can be designed by choosing the routing direction of currents between adjacent inductors. As shown in Figure 3.17(a), a non-alternating

inductor layout makes low-frequency currents between adjacent inductors flow in the same clockwise direction. This results in a destructive magnetic coupling ($M < 0$) between adjacent inductors, and will reduce the effective intrinsic impedance of the ladder. An alternating inductor layout inside the LC-ladder generates a constructive magnetic coupling ($M > 0$) between adjacent inductors. We can achieve the same ladder impedance using a smaller inductor in the alternating LC-ladder as compared to the non-alternating case. In other words, if we utilize an alternating coupling LC-ladder, we not only reduce the spacing between adjacent inductors, but also reduce the size of each individual inductor inside the ladder. The overall ladder layout can be dramatically reduced.

3.7. WDA Schematics and Layout

Based on the WDA noise optimization process explained in Section 3.5, the optimized stage-weights $G_{m,i}|_{optimized}$ at 17 mA were used to implement the WDA. The schematics of the implemented WDA test IC is shown Figure 3.18. The coupling input and output LC-ladders have a design of $50\ \Omega$ intrinsic impedance, and were terminated with variable resistors (R_i and R_o) at the other ends for both the input and output LC-ladders. The input termination resistor R_i was shunted with a bypass capacitor in order to supply gate bias voltage V_{bg1} for stage amplifiers ($G_{m1}, G_{m2}, G_{m3}, G_{m4}$, and G_{m5}). Output bias currents for stage amplifiers were supplied from the RF output through a bias-T, and its termination resistor R_o is in series with a bypass capacitor to ensure the correct output bias voltage.

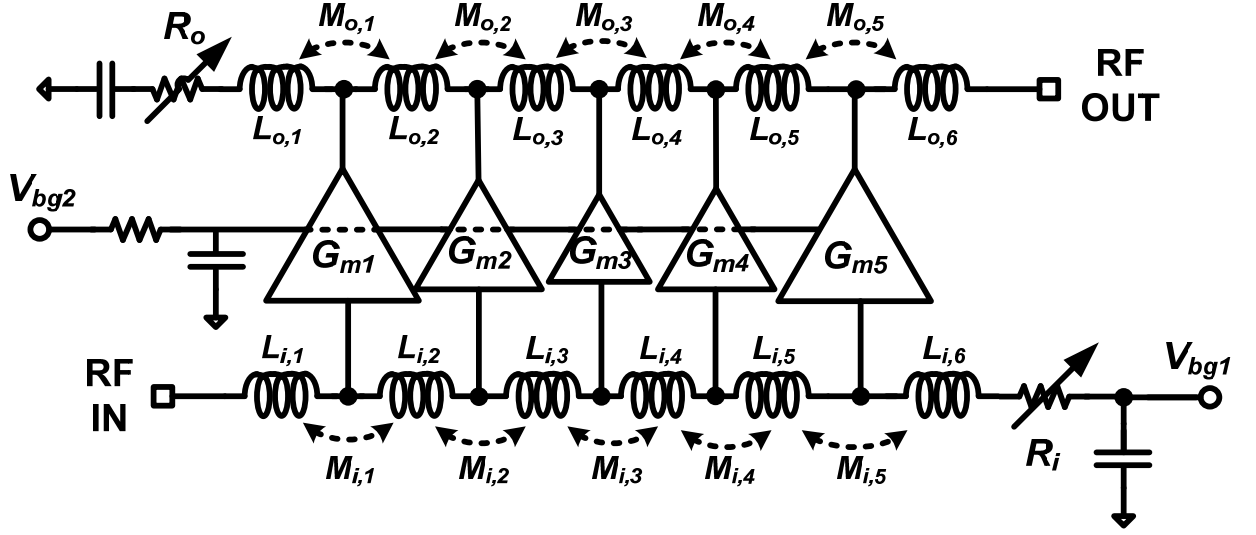


Figure 3.18: Schematics of the WDA

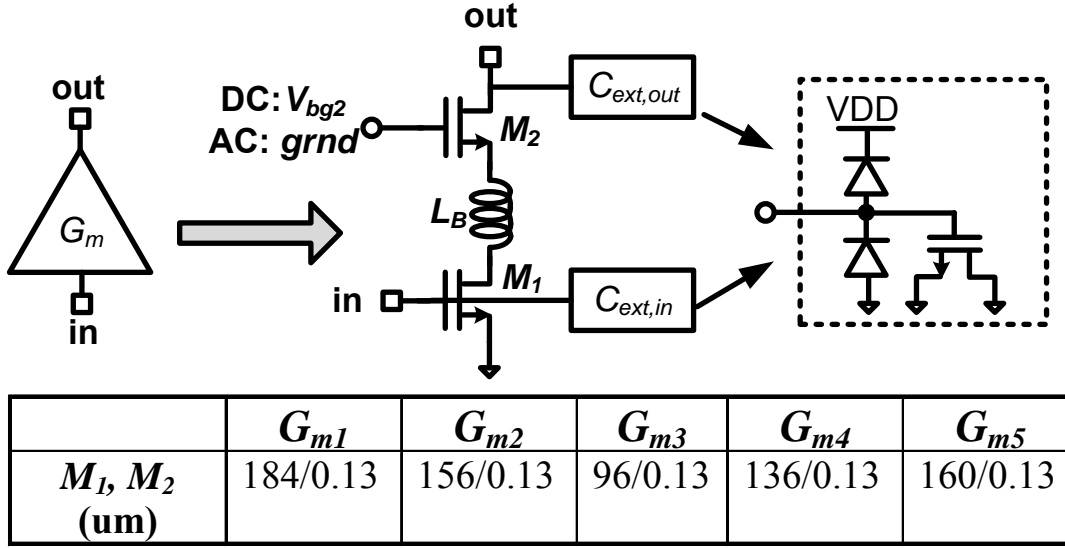


Figure 3.19: Schematics of the intermediate amplifiers and their device sizing

Each weighted stage amplifier is a cascode amplifier with a bandwidth enhancement inductor (L_B) as shown in Figure 3.19 [17]. The transistor sizing inside each stage amplifier is tabulated inside the same figure, and M_1 and M_2 are kept the same to simplify the design process. Other than the first stage G_{m1} , all other stage amplifiers have smaller

weights and smaller input parasitic capacitance. To maintain the uniformity of intrinsic impedance and group delay along the *LC*-ladders, additional parasitic capacitance will need to be supplied. This provides us an opportunity to place ESD diodes in those nodes without sacrificing WDA noise figure and power consumption. Inductors inside the *LC*-ladders present very low impedance at low frequency. As a result, ESD diodes at different stages form an aggregate large ESD protection for the WDA.

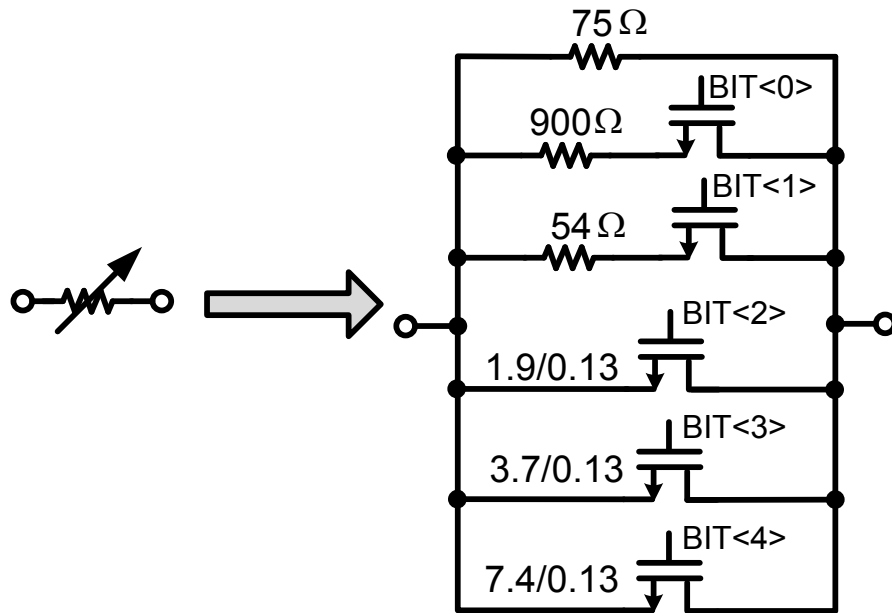


Figure 3.20: Schematics of the variable termination resistors

Poly-resistors in the TSMC 130 nm CMOS process have a large process variation. Since accurate resistance value is necessary for terminating ladders, a variable resistor was designed to cover all the process corners. The schematic of this variable resistor is shown in Figure 3.20. This design presents a tunable resistance from 35 to 70 Ω for a typical process corner up to 11 GHz.

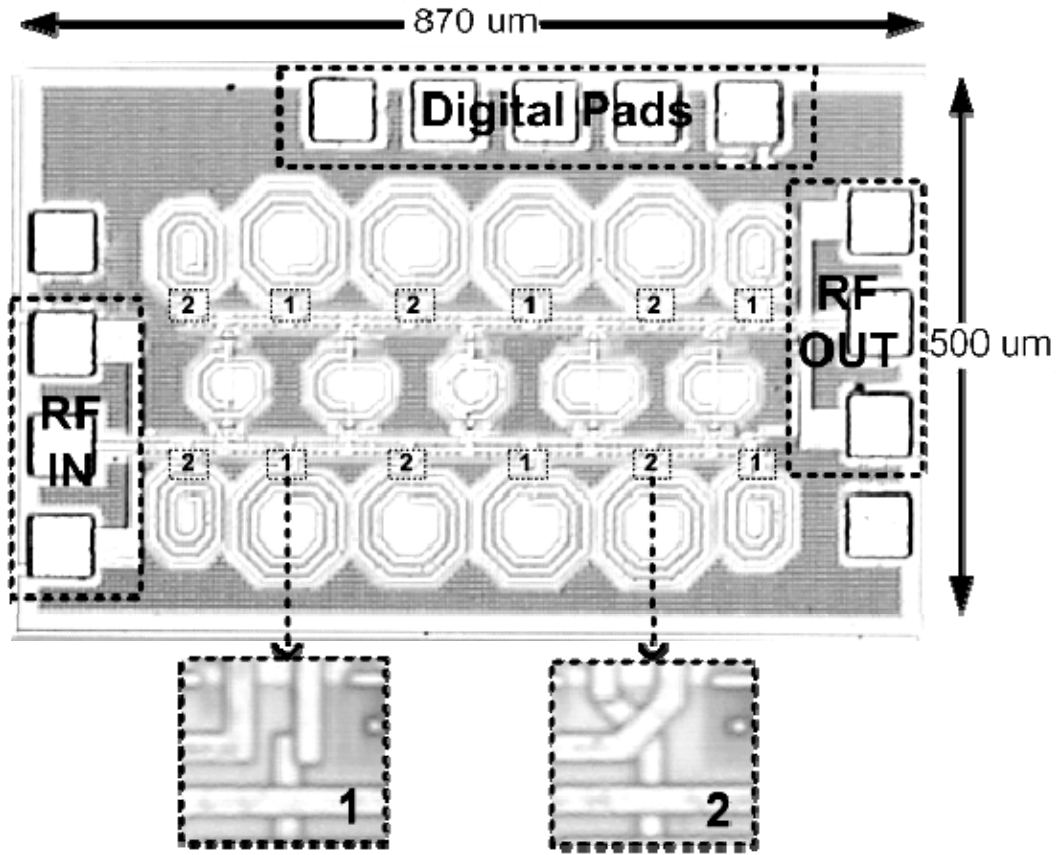


Figure 3.21: Die micrograph of the WDA

The die photo of the IC is shown in Figure 3.21, and this chip is based on TSMC 130 nm CMOS 1P7M process. Chip size is $500 \times 870 \mu m^2$ including all RF and digital pads. EM simulation was carried after initial layout is finished, and an integrated design flow for RF-VLSI, as explained in Appendix 3.1, is used to improve simulation accuracy and the integration between the EM simulator and the regular VLSI verification process. The layout for the alternating coupling LC-ladders is also illustrated in the zoomed-in window inside Figure 3.21.

3.8. WDA Measurement Results

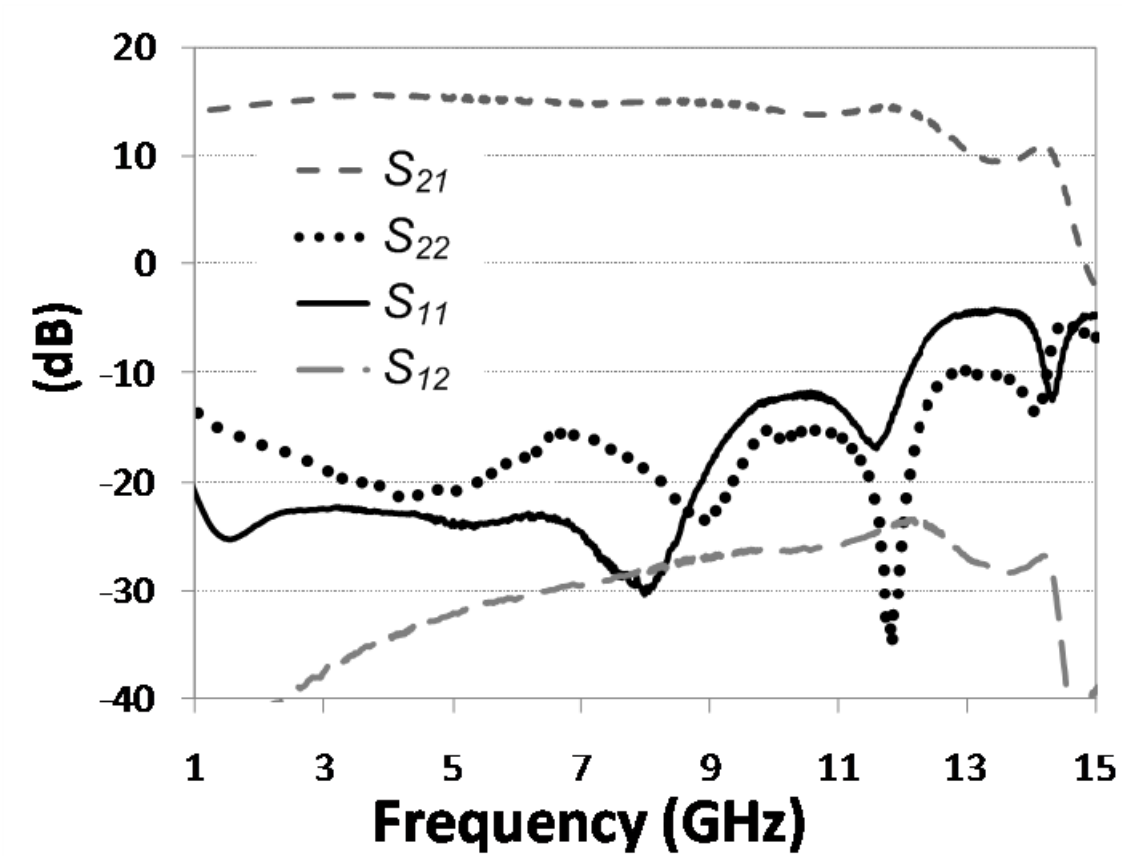


Figure 3.22: S-parameters measurement results

The *WDA* IC was mounted on a printed circuit board with wirebonds to its DC and digital pads. Coplanar probes were used for the RF measurement, with the measurement calibration plane to the middle of the RF pads. Drain bias current was supplied through the output RF probes through a bias-T. Figure 3.22 shows typical S-parameter measurement results of the WDA at 26 mW power consumption. S_{21} ranges from 14 to 16 dB from 1 to 10.6 GHz. S_{11} and S_{22} are better than -12 dB from 1 to 10.6 GHz. S_{12} is better than -23 dB across the same bandwidth.

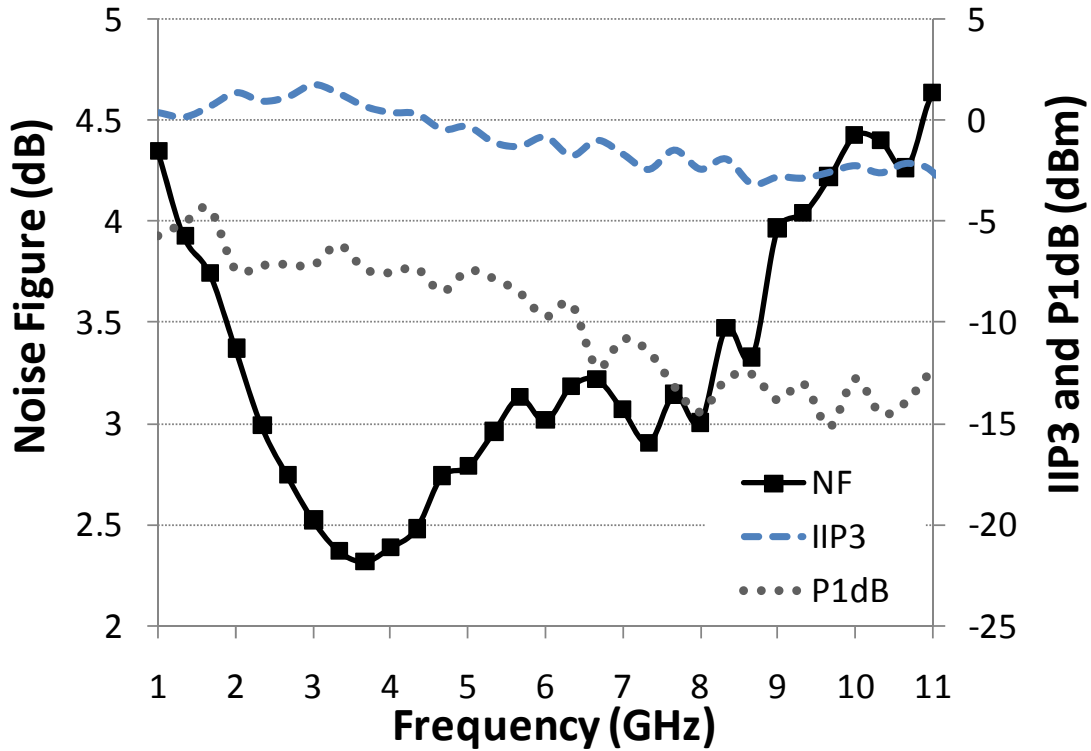


Figure 3.23: Noise figure (NF), input-referred third-order intercept point (IIP3), and input-referred 1 dB gain compression point (P1dB) measurement results at 26 mW

Noise source, cables, connectors, and RF probes have been carefully calibrated for the noise measurement with repeatability errors less than 0.05 dB. As shown in Figure 3.23, the NF of the WDA IC ranges from 2.3 to 4.5 dB and from 1 to 10.6GHz. The rising NF behavior at higher frequency is due to the excessive loss in the LC-ladders due from interconnection skin effects. IIP3 measurement of the WDA IC is better than -3 dBm, and the P1dB is better than -15 dB across the 1 to 10.6 GHz bandwidth. These results are measured at 26 mW power consumption from a 1V power supply.

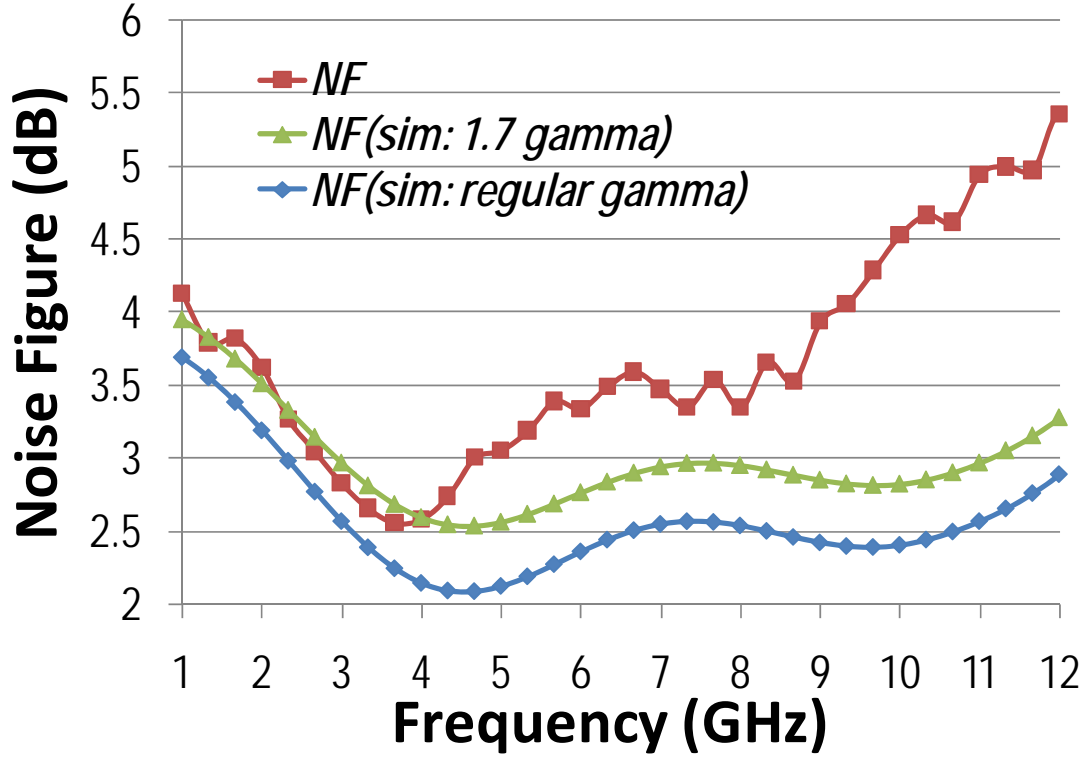


Figure 3.24: Measured noise figure (NF) and simulated NF of different transistor γ at 17 mW power consumption

The measured NF and the simulated NFs for two different transistor γ are plotted in Figure 3.24. The $\gamma = 1.7$ simulated NF is a post-measurement simulation result, which is fitted to the measured NF at lower frequency range. The regular γ simulated NF is the pre-fabrication WDA noise simulation, which utilizes the transistor data provided by the foundry. Inaccuracy in modeling γ results in an almost uniform NF simulation errors across the designed bandwidth. The increasing discrepancy between the measured and simulated NF at higher frequencies suggests that the WDA IC has a higher metal loss due to skin effect than the simulation. The WDA IC also shows a slightly larger ladder group delay than simulation.

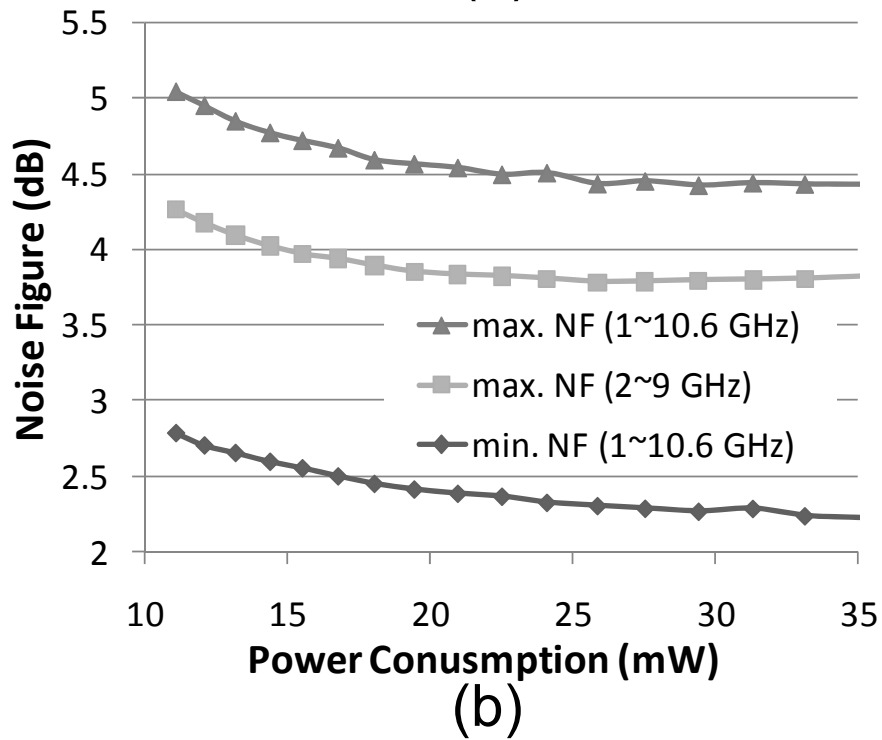
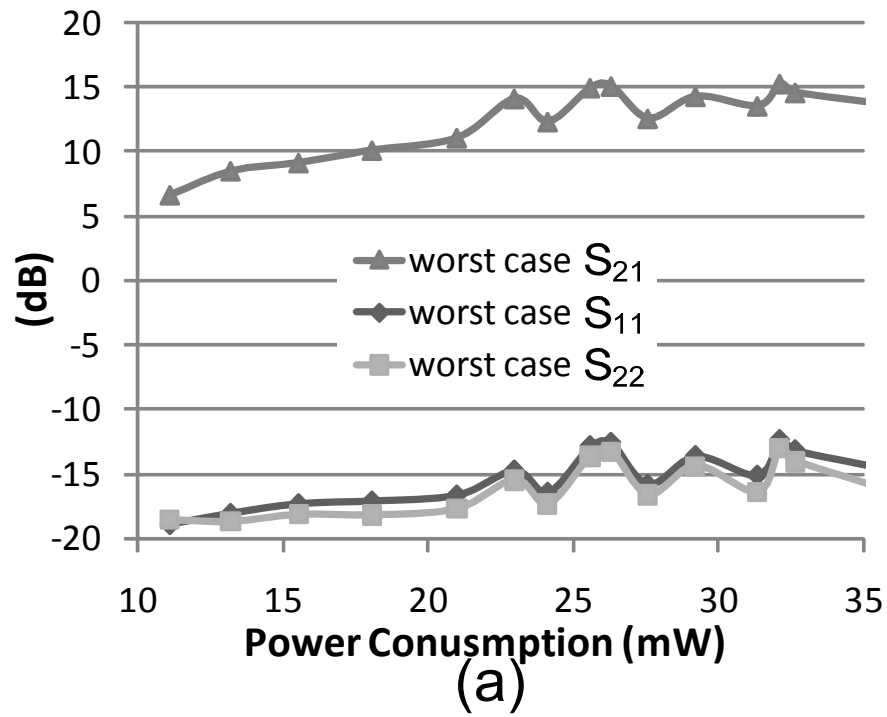


Figure 3.25: Worst-case measured performance versus power consumption

Worst-case WDA measured performance versus power consumption is shown in Figure 3.25. Figure 3.25(a) shows the worst-case input and output matching from 1–10.6 GHz bandwidth for different current consumption. Power consumptions are mainly determined by the gate voltage and the output drain voltage of the stage-amplifiers. Since the gate-to-source (C_{gs}) of a transistor is a function of its bias voltage, the change on the power consumption affects the effective shunt capacitance, and the intrinsic impedance of the ladders. However, we can adjust the termination resistance accordingly by the variable resistors to match the ladders' impedance changes. As a result, the overall input and output matching is good across a wide power consumption range. The worst case S_{21} from 1 to 10.6 GHz is also a strong function of power consumption, and is increasing with power consumption. S_{21} reaches its maximum value at 15 dB.

Worst-case NF is also a strong function of power consumption, and has a opposite tendency to the S_{21} . The NF reaches a plateau when power consumption is greater than 26 mW.

3.9. Summary

In this chapter, we briefly review the Bode-Fano criteria and explain the conventional trade-offs in the design of broadband CMOS LNAs using power-constraint optimization. A distributed amplifier relieves these trade-offs; however, each of its gain stages doesn't contribute equal noise to the output, and its LC-ladders layout takes unnecessarily large area. A weighted distributed amplification concept improves the noise performance of a DA at the same power consumption by utilizing the finite impulse response filtering

property inside the WDA for each noise source. One of the distinct advantages of the WDA topology is its tolerance to I/O parasitic capacitance; ESD protections can also be placed at the smaller weighted gain stages without degrading the performance of a WDA. Alternating coupling LC-ladders are analyzed and utilized in the design to reduce the ladder size. A test IC is implemented in the TSMC 130 nm CMOS 1P7M process and occupies a $500 \times 870 \text{ um}^2$ die area. A 2.3–4.5 dB NF performance is achieved at 23 mW power consumption.

Appendix 3.1: Integrated RF VLSI Design Flow

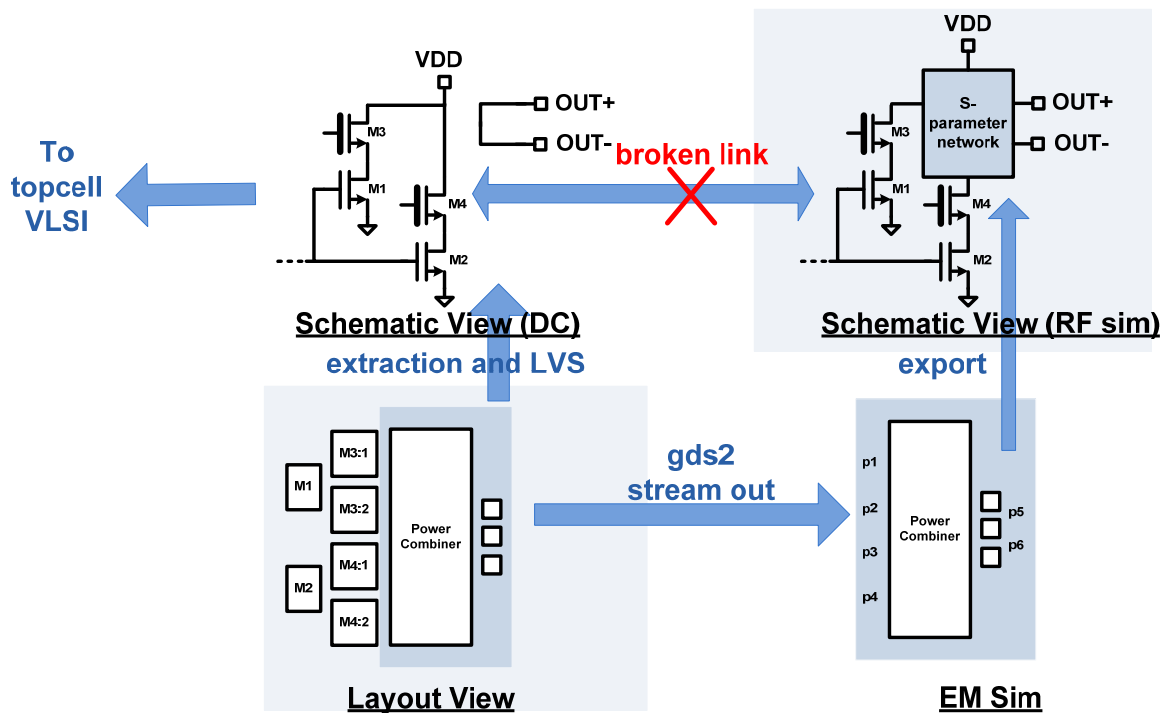


Figure 3.26: Conventional RF IC design flow

Modern VLSI relies on EDA verification tools to ensure the manufacturability of a physical design and the equivalence between the logical and physical design. After a physical layout is verified for its manufacturability and logic equivalence, the EDA tools extract its interconnection parasitic capacitance for post-layout simulation to evaluate the real IC performance after fabrication.

RF IC built on modern VLSI process faces a unique challenge: its signal wavelength is at a comparable dimension to its signal interconnection. Any metal interconnections cannot be viewed as a simple electrical short. Both parasitic inductance and capacitance become significant when the signal frequency increases. In addition, most RFICs contain custom EM structures to improve their performance. These custom EM structures need to be simulated in EM software to calculate the high-frequency behaviors of the structures. Using custom EM structures and simulation results poses two issues for conventional VLSI design flow. First of all, custom EM structures are not integrated with EDA tools, and are usually not provided by the foundries as part of the standard design kit. Additional DC schematic has to be made in order to check layout versus schematics (*LVS*) as shown in Figure 3.26. In addition, we have to use another RF schematic which embeds the EM simulation results to simulate RF performance of the IC. Though EM simulation is very accurate in predicting a structure's EM characteristics, it is not computationally efficient in calculating parasitic capacitance due to the metal interconnections. Using an independent schematic for RF simulation will result in a RF schematic that contains a correct EM simulation result for the EM structure, but underestimates the parasitic capacitance generated from these metal interconnections. Furthermore, since the (DC) schematics used

for LVS are different from the RF schematics, consistency between verification and simulation needs to be checked manually by the RF designers. For complex RF systems, like transceivers, it becomes a non-trivial job to keep verification and simulation consistency. Also, since the DC schematic/netlist is used at the system *topcell*, it becomes impossible to simulate RF performance on the system level.

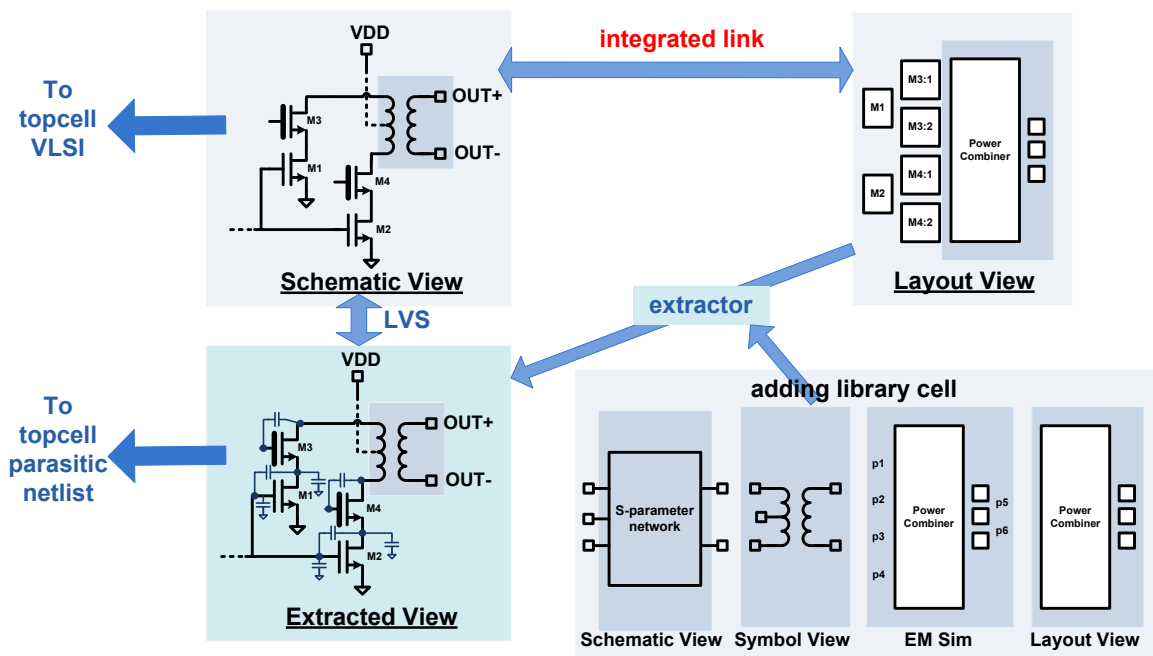


Figure 3.27: An integrated RF VLSI design flow

An integrated RF VLSI design flow assumes that un-EM-simulated parasitic capacitance is limited to a local area in terms of signal wavelength, and resorts to the parasitic extractor to calculate these parasitics. In other words, these parasitic capacitors are assumed to be lumped locally. RF designers have to decide which part of the RFIC needs to be EM simulated, while the rest of the metal interconnections are left for parasitic extraction.

An integrated design flow starts from making custom EM structures into custom library cells which contain a symbol view, a layout view, and a schematic view that contains the EM-simulation results of the structures. Designers need to manually check the equivalence of the EM structure between its symbol, layout, and schematic views. Since manual consistency checking on the library cell level is considered manageable, this reduces the chance of errors in consistency checking. Once a library-cell is made, it is then integrated into the parasitic extractor of the EDA design environments.

Verification process starts with parasitic extraction that extracts the layout into an extracted schematic and netlist. Since the custom EM structures are built into the extractor, the extracted netlist will contain the correct EM cell, as well as the interconnection parasitic capacitance. The schematic view is the only circuits for both DC and RF simulations which excludes the parasitic capacitance from the interconnections, and LVS checking uses this schematic view to verify the consistency between schematic and the layout view. The extracted netlist contains both the custom EM structures as well as the extracted parasitic, and can be used for topcell system RF-VLSI co-simulation. In this manner, RF-VLSI verification is consistent with its simulation process on the topcell, and both EM effects and interconnection parasitics have been considered in the simulation.

Chapter 4: Concurrent Octa-Core RF Receiver Architecture

The concurrent use of several RF channels for wireless communication will increase the effective communication bandwidth, and boost the transmission data rate. In this chapter, we will propose dynamically scalable concurrent communication, which divides the 7.5 GHz bandwidth of the 3.1–10.6 GHz unlicensed band into seven concurrent channels. The concurrent use of these channels results in multi-GHz analog bandwidth to support multi-Gbps wireless communication. A RF multi-core RF system architecture is then proposed and implemented to verify the concept. Compared with previous works using this band, the proposed architecture has better spectrum efficiency. In addition, the multi-core RF system architecture is well-aligned with the trend of multi-core digital processing in high-performance applications, where the best performance is achieved with a larger number of parallel cores instead of a single higher speed processor.

In Section 4.1, we will first introduce the applications of wireless multi-Gbps communication. In the U.S., there are two unlicensed bands which have multi-GHz bandwidth to support wireless communication at multi-Gbps data rate, we will discuss the pros and cons of the two bands. A short discussion of previous works using the 3.1–10.6 GHz band will also be covered. In Section 4.2, we introduce the octa-core RF receiver architecture, explain block diagrams, and discuss its IC implementation. A receiver

prototype is implemented in a 130 nm CMOS process, and has been measured to verify the design concept. A summary of this chapter will be provided in Section 4.3.

4.1 Introduction

4.1.1 Wireless Multi-Gbps Communication

Many applications require or benefit from high data rates far exceeding the capability of existing wireless technology. One of such example is the wireless transmission of uncompressed high definition video signals. The direct sending of uncompressed video signals greatly reduces power overhead for encoding and decoding video. Set-up boxes, Blu-Ray DVD players, and digital video cameras will be the beneficiaries of this technology. In general, the need for bandwidth is insatiable, much like the demand for CPU speed, static and dynamic RAM, flash memory, and external hard disk capacity.

To establish a high-speed wireless link, we need to first allocate communication channels. Each channel has a maximum achievable data rate, known as channel capacity C . Channel capacity is related to the bandwidth of the channel, BW , and the signal-to-noise ratio, SNR , in the following manner [49]:

$$C = BW \cdot \log_2(1 + SNR), \quad (4.1.1)$$

which shows that we can either use a large BW or a higher SNR to achieve a high communication data rate. However, based on the theoretic studies of digital communications [50], to achieve the same data rate using a smaller BW means to resort to

more efficient use of the spectrum. As an example, to extract 1 Gbps from 100 MHz of bandwidth channel obviously requires 10 bits per Hz, but only 1 bit per Hz from a 1 GHz bandwidth channel, and low-order constellations can be used to transmit and receive the data. The narrow-bandwidth system (100 MHz BW in this example), on the other hand, must use sophisticated signal modulation, often placing stringent demands on phase noise and power amplifier linearity (particularly for OFDM), and this translates into a system with less overall sensitivity. Much energy must be consumed in the baseband of the narrow-band systems to provide FFT and equalization functionality, which will end up consuming more energy per bit than the wide-bandwidth solution. Hence, a large bandwidth can reduce the complexity of communication system and reduce overall power consumption.

In the U.S., at the time this thesis is written, the Federal Communication Commission (FCC) has allocated three unlicensed bands with more than 1 GHz bandwidth. They are the 7.5 GHz bandwidth in the 3.1–10.6 GHz [51], the 7 GHz bandwidth in the 57–64 GHz (60 GHz band), and the 3 GHz bandwidth in the 92–95 GHz (90 GHz band) [52]. From RF circuits' point of view, the implementation RF transceiver in 90 GHz band is similar to that in the 60 GHz band except at a higher center frequency. Between 60 GHz and 90 GHz, the 60 GHz attracts more research attention by far. In Section 4.1.2, we will compare the 3.1–10.6 GHz and the 60 GHz band.

4.1.2 Comparisons between the 3.1–10.6 GHz and the 60 GHz Band

Given an unlicensed 7.5 GHz bandwidth, the 3.1–10.6 GHz band has received less attention in the race of multi-gigabit wireless communication for two major reasons: Firstly, this band's wide baseband bandwidth over center frequency makes the conventional single-carrier-based analog frequency-translation scheme ineffective. Secondly, the low equivalent isotropically radiated power (EIRP) limit enforced by the FCC makes the band unattractive to high performance applications. However, the 3.1–10.6 GHz RF signal has on the average 20 dB less channel path loss compared to the 60 GHz band to justify its low EIRP. In addition, at lower microwave frequency, RF signal would more easily penetrate through or diffract around obstacles along the wireless link, which makes non-line-of-sight communication possible. Furthermore, the 60 GHz band is much closer to the transistor f_T compared to its 3.1–10.6 GHz band, and this implies a more inefficient power generation and amplification for the RF frontend circuitry. Also, RF packaging in the 60 GHz will be more difficult due to its high frequency.

4.1.3 Previous Works using the 3.1–10.6 GHz Band

The unlicensed 3.1–10.6 GHz band has a 7.5 GHz total BW and can be used for short-distance multi-gigabit wireless communication. Previous works utilizing this band fall under two major categories: time-domain impulse-based [53][54][55], or WiMedia's MB-OFDM compliant (frequency-hopping based) [56][57]. The impulse-based method has low spectral efficiency and is susceptible to inter symbol interference (ISI) for high data-rate

communication due to the relatively large multipath delay spreads over the pulse period, as shown in Figure 4.1(a). Equalization of the received impulse signals with ISI will be difficult to be implemented due to multi-GHz clocking speed of the digital signal processor. As a result, it is difficult to establish a reliable Gbps wireless link using the impulse-based approach.

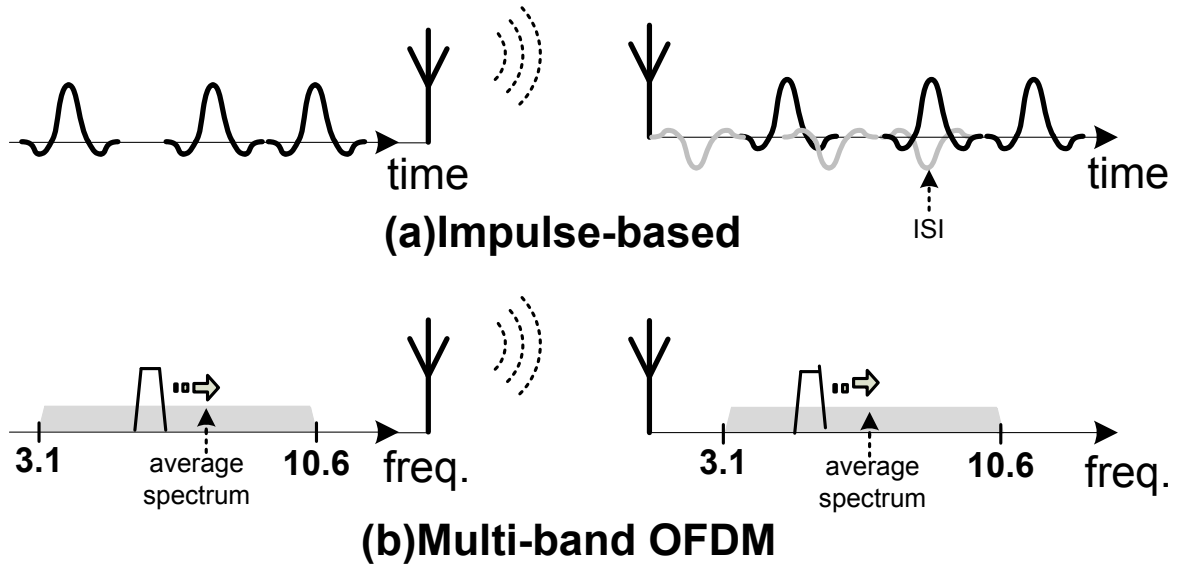


Figure 4.1: Previous works on the 3.1–10.6 GHz band: (a) Impulse-band and (b) Frequency-hopping based

The MB-OFDM-compliant approach is essentially a diversity-improved narrow-band (frequency-hopping) method that utilizes a 528 MHz band out of the available 7.5 GHz RF spectrum at a given time window, as shown in Figure 4.1(b). The average spectrum over long measurement time window will follow the FCC's radiation emission regulation on ultrawide-band signal. However, since only 528 MHz is effectively used for wireless

communication, it requires a complex modulation scheme to achieve a multi-Gbps data rate. In addition, the requirement of frequency hopping will increase circuit overheads in the hopping LO implementation, and result in high power consumption.

We propose to divide the 3.1-10.6GHz into several RF channels and use them independently and concurrently. We will discuss this communication scheme and its RF receiver implementations.

4.2 A 3.1–10.6 GHz Octa-Core Receiver

In this section, we will discuss the dynamically scalable concurrent communication which fully utilizes the 3.1–10.6 GHz spectrum. A CMOS octa-core RF receiver IC will be studied, implemented, and measured to realize concept.

4.2.1 System Architecture

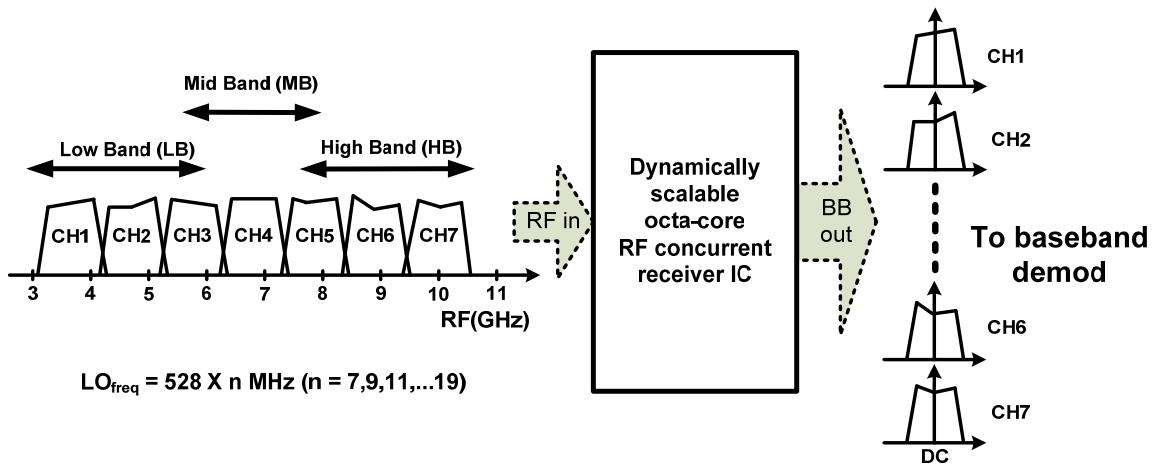


Figure 4.2: Dynamically scalable concurrent communication

Figure 4.2 illustrates the basic concept of the proposed dynamically scalable concurrent communication, and the role of the octa-core RF concurrent receiver IC. This approach divides the 3.1–10.6 GHz unlicensed spectrum into seven channels, and uses a variable number of the channels ranging from one to all seven, depending on the channels' availability and the needed data rates. The center frequencies (LO_{freq}) of these RF channels are $LO_{freq} = 528 \times n \text{ MHz}$, where $n = 7, 9, 11, \dots, 19$. These RF channels' BW is identically 1.056 GHz. Using the same baseband bandwidth for different RF channels can reduce design complexity of the baseband signal processor. The role of the octa-core RF receiver in this scheme is to concurrently down-convert selected RF channels to baseband. Each of the concurrent channels has a reduced baseband BW corresponding to a 1.056 GHz Nyquist rate, as opposed to the full 7.5 GHz BW. This greatly reduces both the clocking rate and the dynamic range requirement of the baseband signal processor. This approach is well-aligned with the trends of multi-core digital processing in high-performance applications, where the best performance is achieved using a larger number of parallel cores instead of a single higher speed processor. The energy spent in communicating any single bit can thus be minimized.

The system architecture of the proposed octa-core RF receiver IC is shown in Figure 4.3. The system consists of a main RF amplification common part and eight independently controlled down-conversion cores. The main RF amplification common part consists of a weighted distributed amplifier (WDA), a global RF buffer, and an RF balun followed by a signal distribution line to feed wideband signals to the eight down-conversion cores. The function of the main RF amplification common part is to amplify the broadband RF input

signals with little added thermal noise, so the down-conversion core will not degrade system noise figures.

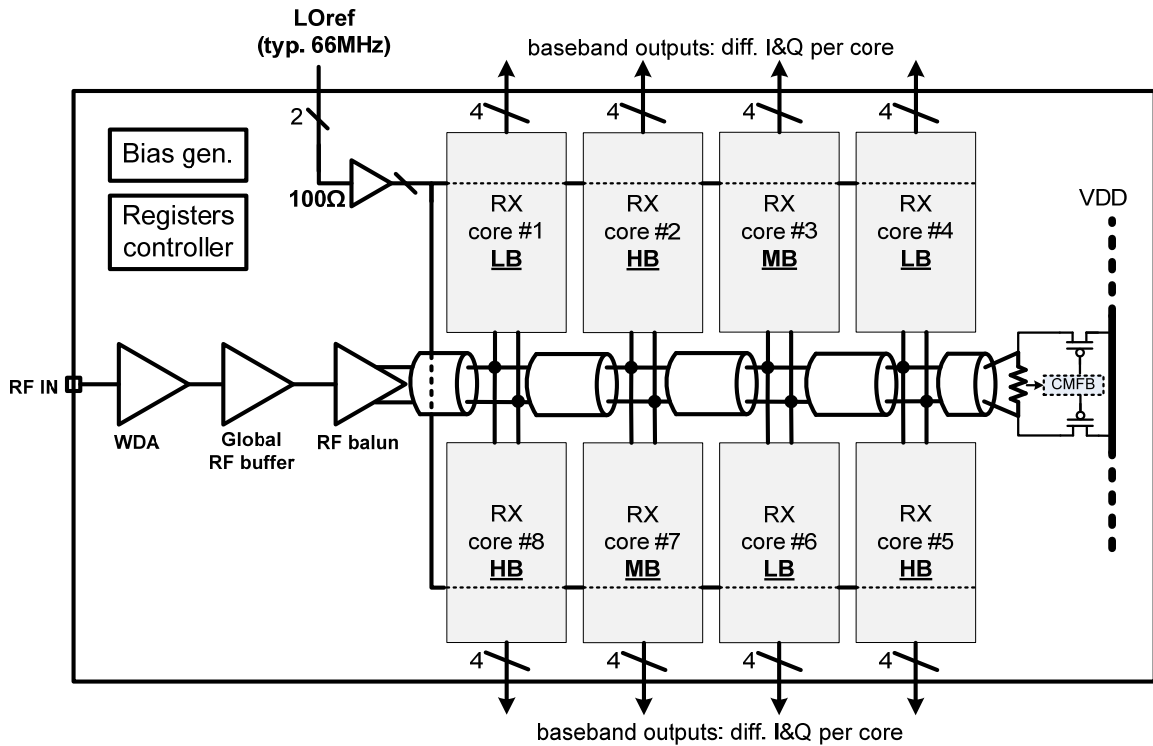


Figure 4.3: System architecture of proposed octa-core RF receiver

Each down-conversion core is comprised of a frequency synthesis block, I&Q down-conversion mixers, and the baseband variable gain buffers. The function of each down-conversion core is to down-convert a particular RF channel into corresponding I&Q basebands so the out-of-band signals and the LO leakages will be filtered out by the baseband filter. The MOS varactors of the LC-voltage control oscillator (LC-VCO) inside each frequency synthesis block have limited tunability. To cover all seven RF channels, three different versions of the down-conversion cores, i.e., low band (LB), mid band (MB),

and high band (HB), are designed. At a typical process corner, the LB core can be programmed to down-convert any of the first three RF channels. The MB core covers the third to the fifth RF channels, and the HB core processes the fifth to the seventh channels. This frequency plan is shown in Figure 4.2. Inside each chip, there are three LB cores, two MB cores, and three HB cores. The overlapped frequency plan combined with the extra (eighth) core provides necessary redundancy to cover all seven bands in the presence of large systematic process variations in the VCO center frequencies.

There are two major design challenges of this architecture: core-to-core interference and power consumption. At the worst case, there will be seven cores running on a silicon die; severe core-to-core interference can make this architecture useless. In addition, each of the seven cores has its own down-conversion signal paths and frequency synthesis. Power consumption can be excessive if the system and circuits are not designed properly.

Core-to-core interference is alleviated first by careful frequency planning. By choosing $LO_{freq} = n \times 528 \text{ MHz}$ ($n = 7, 9, 11, \dots, 19$), avoidance of third-order spurious harmonic mixing from the lowest to the highest channel is guaranteed. In addition, the lack of simple fractional relationship between any two of different cores' LO frequencies prevents VCOs from pulling and interlocking. Also, careful interleaving placement of different cores on chip (as shown in Figure 4.3, the sequence of core placement from the left to the right on the lower row is: HB, MB, LB, and HB) and the dynamic allocation of LO frequencies of the cores in real time ensures maximum physical distance between cores with adjacent LO frequencies. Furthermore, each core is surrounded by wide guard rings

with strongly over-damped supply bypass. This further attenuates substrate couplings and supply/ground network perturbations.

Reduced system power consumption is achieved on both system and circuit levels. The use of one PLL per core minimizes the routing distance of the high-frequency LO while only the low frequency system reference (typically 66 MHz) is routed across the whole chip. This arrangement reduces the total system fCV^2 power consumption due to LO routing, which dominates significant power consumption in a typical RF receiver. Inside each PLL, divider chains use true-single-phase-clocking (TSPC) logic [58] to reduce both static and dynamic current consumption, except for the first two current-mode logic (CML) necessary prescalers for high frequency operation. Furthermore, bias currents or voltages of most circuit blocks can be adjusted by digital controlling. This allows the receiver to use only necessary power. This also allows all blocks and down-conversion cores to turn into sleep-mode when not used.

The chip has an on-chip bias generation and distribution, so only a reference current is required off-chip. This chip also has a total 1092 bit serial digital controller to independently program the functional settings and bias of each block.

4.2.2 RF Common Part: LNA, RF Buffers, and RF Distribution

Network

Figure 4.4 shows the schematics of the RF common part. The WDA has been discussed in Chapter 3, and its output is connected to a diode-connected PMOS with a 50 Ω termination resistor between the PMOS's drain and gate. The gate of the PMOS is DC-

bypassed. This setup provides both the DC-bias current, and a high-frequency $50\ \Omega$ output load for the WDA. The RF buffer amplifier is a cascode amplifier with an inductor shunt peaking load to provide a broadband response up to 11 GHz. The RF balun is a one-side-AC-grounded differential amplifier. The common-mode rejection of a regular differential amplifier provides a rough single-mode-to-differential-mode signal conversion. The differential operation of later gain stages provides additional common-mode signal rejections. The output current from the RF balun is feeding to a differential transmission line that distributes the amplified broadband signal to the eight down-conversion cores. The transmission line is terminated with a resistor pair with common-mode-feedback circuitry to provide both the correct DC bias voltage for the eight cores, DC bias current for the RF balun, and the right RF impedance for the transmission line. Bias currents/voltages of the RF buffer and RF balun can be adjusted by digital programming.

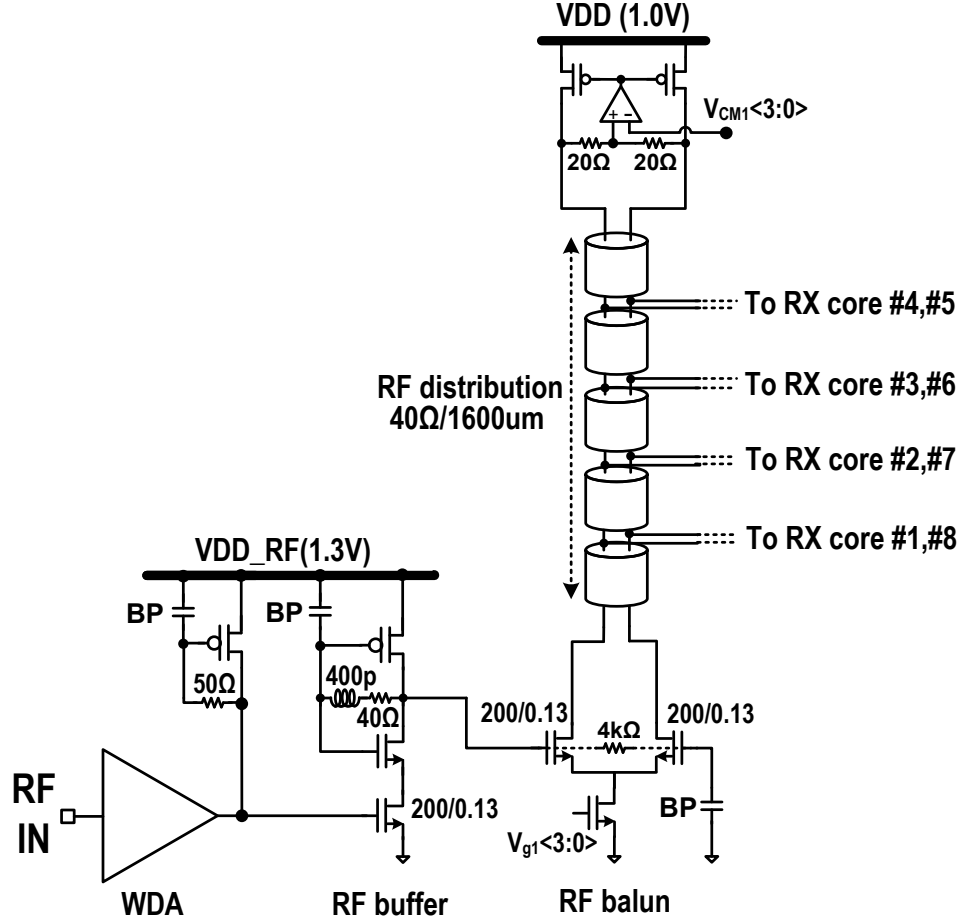


Figure 4.4: Schematics of the RF common part

4.2.3 Downconversion Core: PLL, Mixers, and BB Buffers

Figure 4.5 shows the block diagram of a down-conversion core, which includes a complete I&Q down-conversion signal path and frequency synthesis. The down-conversion signal path starts with a local RF-buffer which amplifies the broad-band signals from the differential transmission line to the inputs of I&Q mixers. The local RF-buffer presents capacitive input impedance, which is absorbed into the differential transmission-line. I&Q mixers use Gilbert-type current commuting double-balanced topology, and the

I&Q LO signals are provided by the in-core PLL. The down-converted signal is amplified by a baseband VGA. A two-step buffer drives a typical differential $100\ \Omega$ load.

Inside each core, frequency synthesis is accomplished by an integer- N phase-locked-loop (PLL). This PLL is comprised of a LC-VCO, two cascading CML div-two prescalers, a CML-to-TSPC converter, a TSPC div-two divider, a modular programmable div- N divider [59], a phase-frequency detector (PFD), a charge-pump, and a second-order low-pass filter connecting in closed loop [48]. There are LB, MB, and HB versions of LC-VCO for the system to cover all of the seven channels. The programmable div- N divider provides dividing ratio from 4 to 31, so the overall programmable divider ratio ranges from 32 to 248, with a step size equal to 8. The PLL LO reference is 66 MHz, and the PLL can generate the required center frequency for the RF channels with proper divider setups. The low-pass filter has a typical bandwidth of 5 MHz, which is large enough to suppress VCO phase noise while small enough for the PLL to remain stable. The output of the LC-VCO is buffered to a RC-CR quadrature filter, and is furthered buffered to drive the LO ports of I&Q mixers.

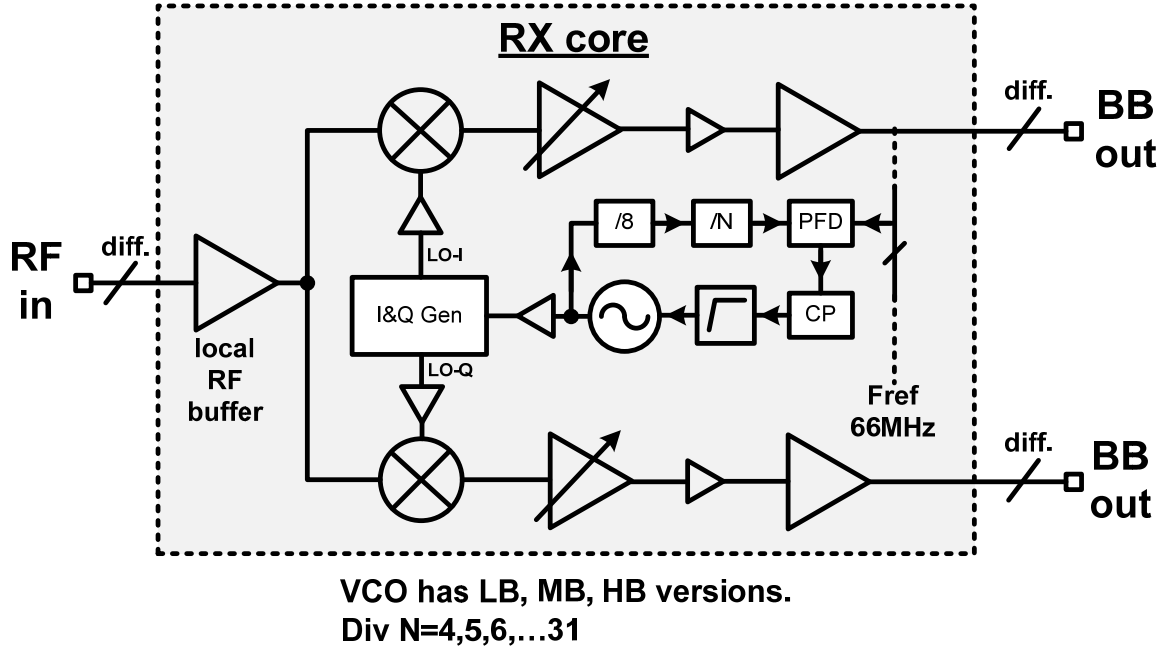


Figure 4.5: Block diagram of the down-conversion core

The octa-core receiver is implemented in a 130 nm CMOS process with seven metal layers. Figure 4.6 shows the die micrograph of the octa-core RF receiver IC, which occupies $1.3 \times 2.7 \text{ mm}^2$. The chip consumes 1 mW in sleep mode. In normal operation, the WDA and the RF buffer consume 29 mA from a 1.3 V supply. The RF balun consumes 21 mA from a 1 V supply. Average current consumption from each down-conversion core excluding the output buffers is 30 mA, with a 55 mA maximum gain current consumption from a 1 V supply. Each differential BB output buffers consumes 5–25 mW from a 1 V supply. Typical total power of the chip with N cores running is $(62 + 40 \times N) \text{ mW}$.

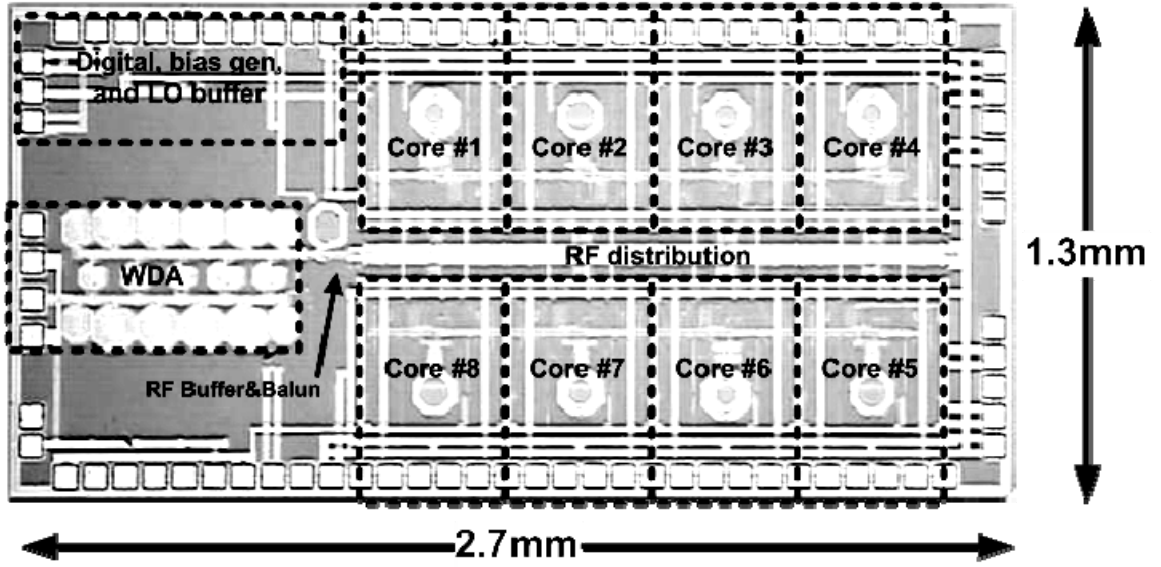


Figure 4.6: Chip micrograph of the octa-core receiver IC

4.2.4 Experimental Results

A printed circuit board (PCB) is designed using a Duroid substrate of a 0.254 mm thickness for prototyping. The PCB provides the traces for the DC supplies, reference signal, digital signals, and differential baseband outputs. All signal inputs and outputs are fed with SMA connectors. The PCB is attached on a gold-plated brass board. Then, through a pre-cut aperture of the PCB, the chip is mounted directly on the brass board using silver epoxy in order to provide good substrate grounding and heatsink. The chip ground pads are wire-bonded directly to the brass board, and the remaining pads are wire-bonded to the PCB traces.

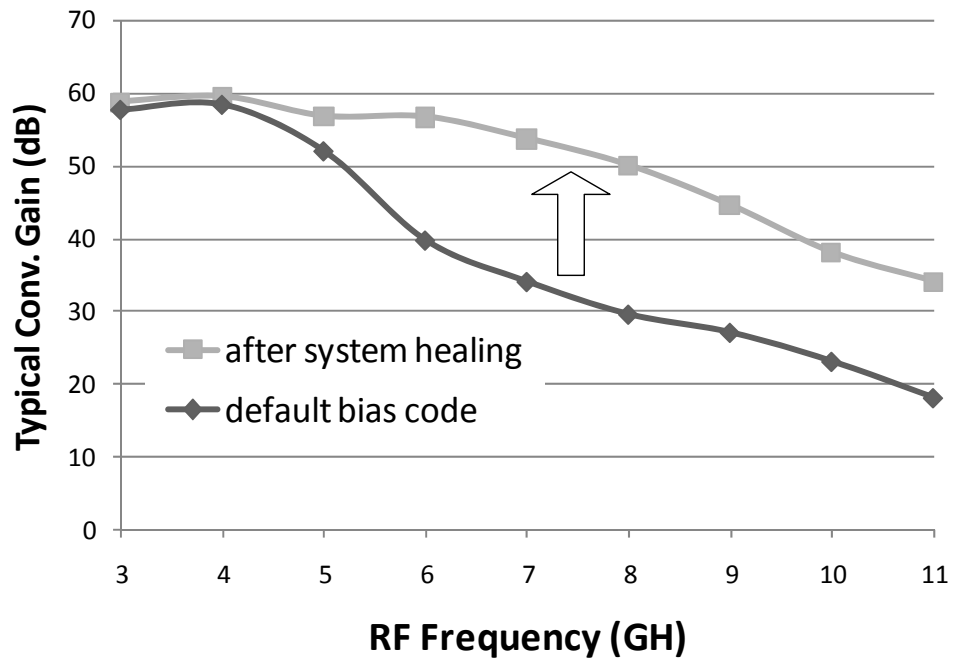
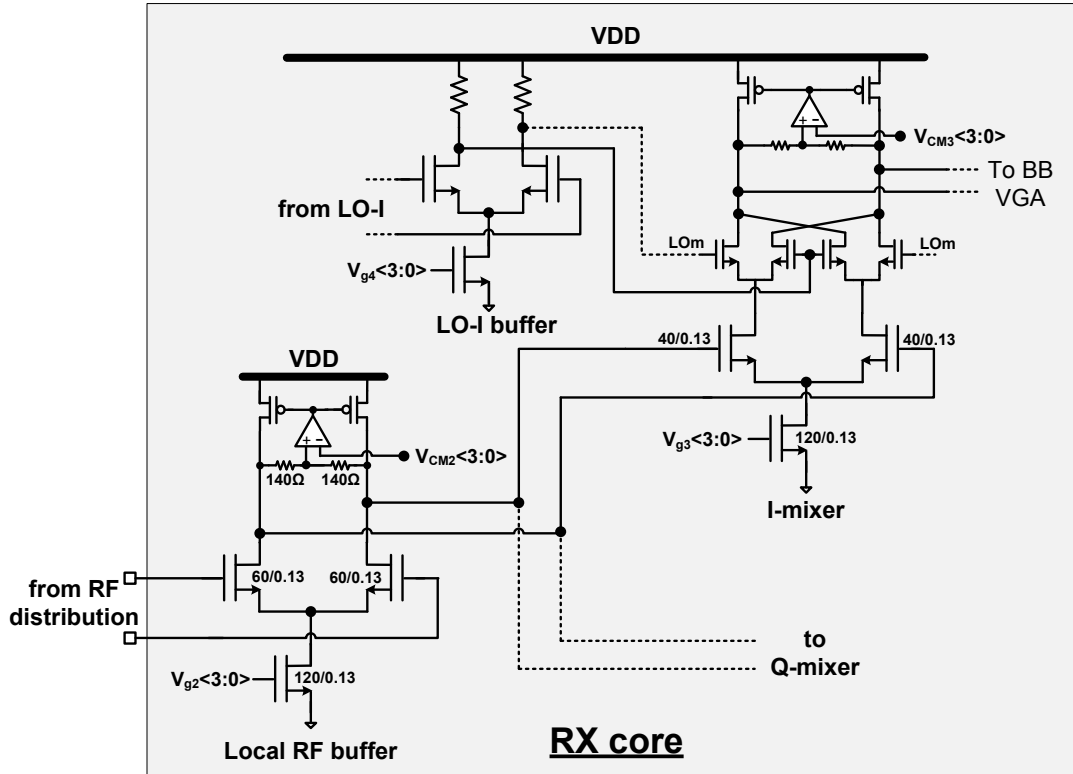


Figure 4.7: Schematics of mixer and RF/LO-I buffers inside each RX core, and the results of system healing at a typical VGA gain setting

Each analog block has an independently controllable bias current with an adjustable common-mode voltage. This allows each block not only to use necessary power, but also to calibrate the circuits for process-related bias variations. As an example, the schematics of one of the mixers and its supporting circuitry, shown in Figure 4.7, are used to explain the process. The operation point of these elements can be dynamically adjusted during the healing phase. The operation points are optimized for the best performance. A typical healing improvement in the conversion gain is also shown in Figure 4.7. Under the same VGA settings, conversion gain is improved by 12 dB on the average after the system calibration.

Figure 4.8 shows the aggregate measured maximum conversion gain performance of the LB, MB, and HB over the 3–10.6 GHz range. The receiver achieves an S_{11} of better than -15 dB up to 11 GHz. The decreasing conversion gain with increasing RF frequency is attributed mainly to the decreasing LO signal level at higher frequency at the LO-port of the RF mixers. This decreases the effective conversion gain of the RF-mixers. Figure 4.9 shows the system noise figure performance of the receiver at maximum gain setting, which ranges from 2.6 to 11 dB. The system NF is dominated by the front-end WDA for lower frequency range. At higher frequencies, the effective conversion gain of the mixer is lowered by the reducing LO signal, and NF from later stages become dominant. Figure 4.9 also shows the minimum-gain system IIP3 which is better than -9 dBm across all frequency ranges. At minimum gain system setting, system IIP3 is limited by the voltage swing at the input node of the global RF buffer right after the WDA.

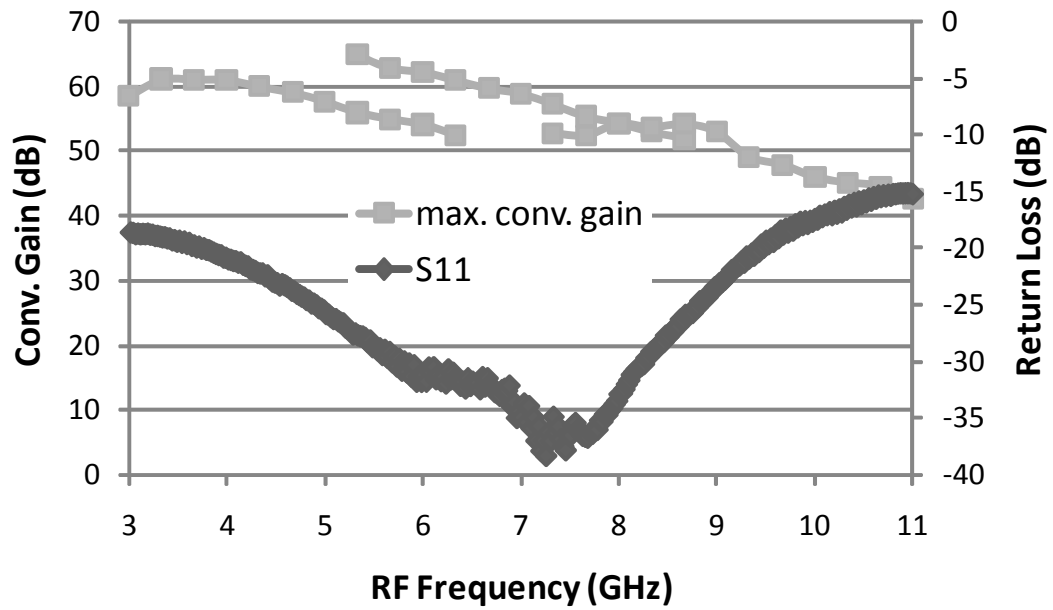


Figure 4.8: Measured receiver maximum conversion gain and S_{11}

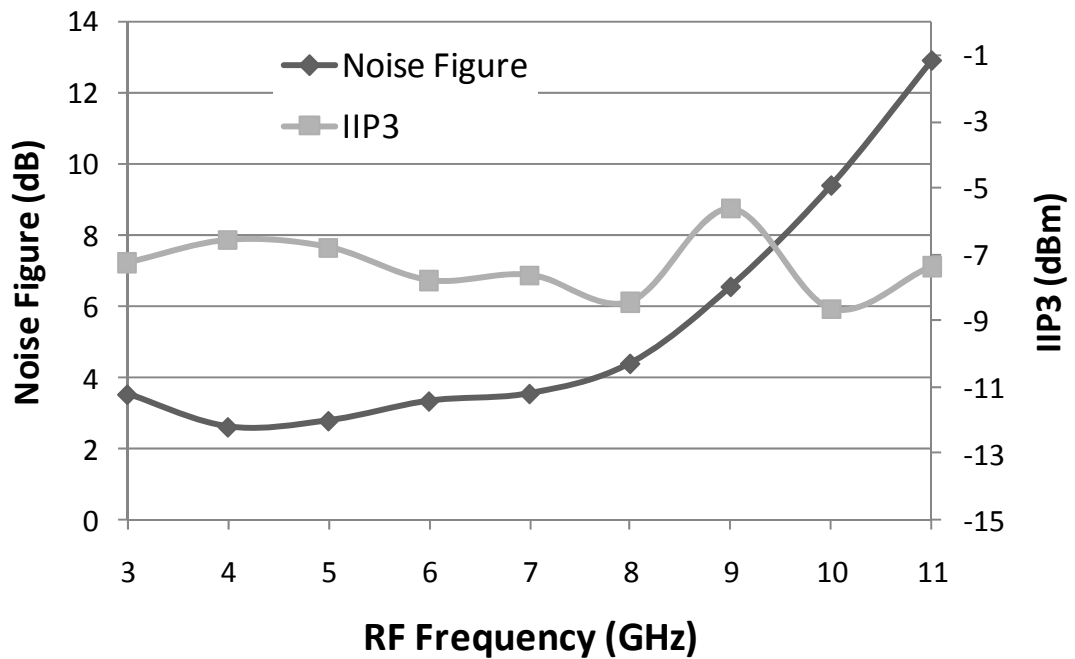


Figure 4.9: Measured receiver system noise figure and IIP3

Multi-core performance is measured with seven active cores running concurrently. No LO pulling is observed when all seven cores are configured to different channel LOs. Off-chip baseband filter with 400 MHz 3 dB BW are used for baseband filtering in the measurement. Cross-band rejection is defined as the unwanted signal power reduction as compared to an in-channel RF signal of the same power, and is an aggregate response of the receiver and the off-chip baseband filter. As shown in Figure 4.10, a better than -36 dBc concurrent cross-band rejection is achieved. Figure 4.11 shows a worst-case -26 dBc LO spur across all cores, and a worst-case -64 dBm concurrent core-to-core adjacent LO leakage at the outputs. This receiver achieves typical 50 mW/GHz power consumption over signal bandwidth when all cores are working concurrently.

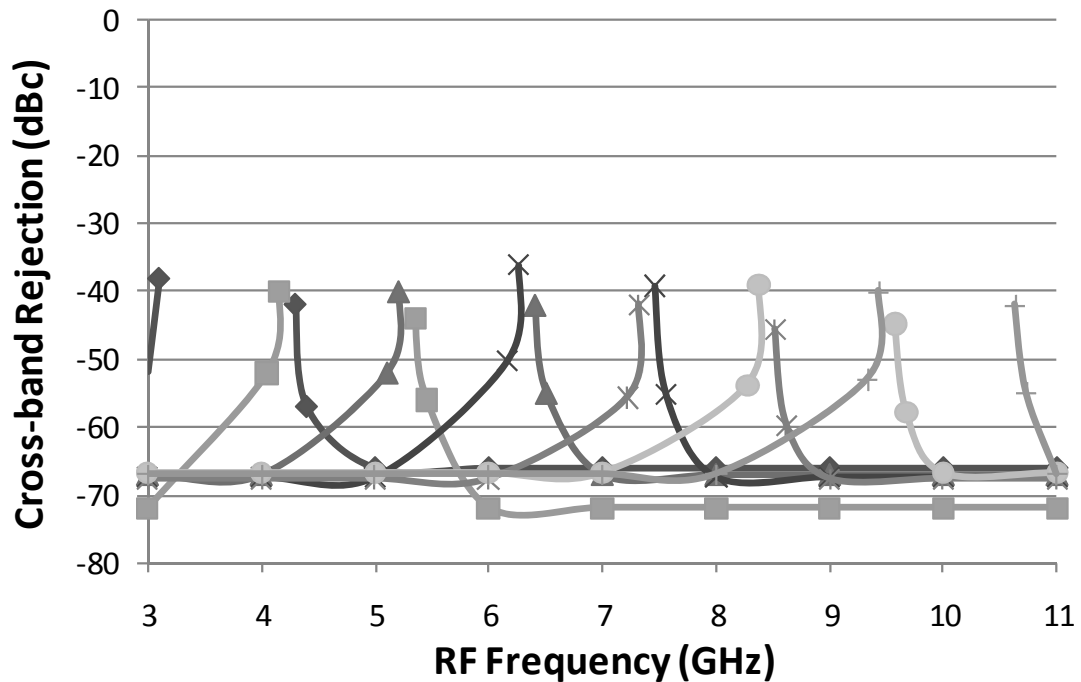


Figure 4.10: Measured cross-band rejection

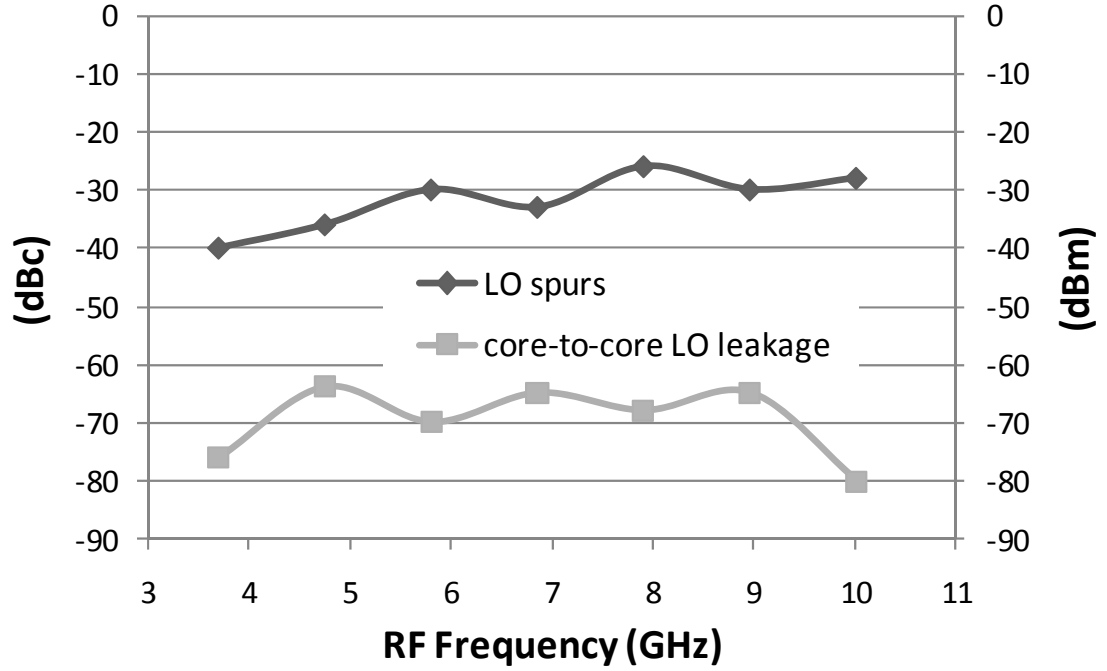


Figure 4.11: Measured LO spurs and core-to-core LO leakage

We can calculate a wireless link capacity based on this concurrent receiver. First, we assume the transmitter transmits at FCC's spectrum mask for UWB band, and both the TX and RX antennas have a 0 dBi antenna gain. Based on the measured receiver system noise figure, we will be able to calculate the SNR_i for the i -th channel using Friis' equation:

$$SNR_i = \frac{P_{ti} G_t G_r \left(\frac{c}{4\pi f D_{tr}} \right)^2}{BW_i \times kT \times NF_i}. \quad (4.2.1)$$

Here, P_{ti} is the transmitted power of i -th channel, G_t is the antenna gain on the transmit side, G_r is the antenna gain on the receive side, c is the speed of light, f is the signal frequency, D_{rt} is the distance between the transmitter and receiver, BW_i is the bandwidth

of the i -th channel, k is Boltzmann's constant, T is the receiver's temperature, and NF_i is the NF of the i -th channel. And the concurrent link capacity can be calculated to be:

$$C = \sum_{i=1}^7 BW_i \cdot \log_2(1 + SNR_i). \quad (4.2.2)$$

The link capacity for different TX-RX distance is calculated and plotted in Figure 4.12. The wireless link based the octa-core RF receiver achieves a theoretic 16 Gbps channel limit at a five meter RX-TX distance. The measured performance summary is shown in Table 4.1.

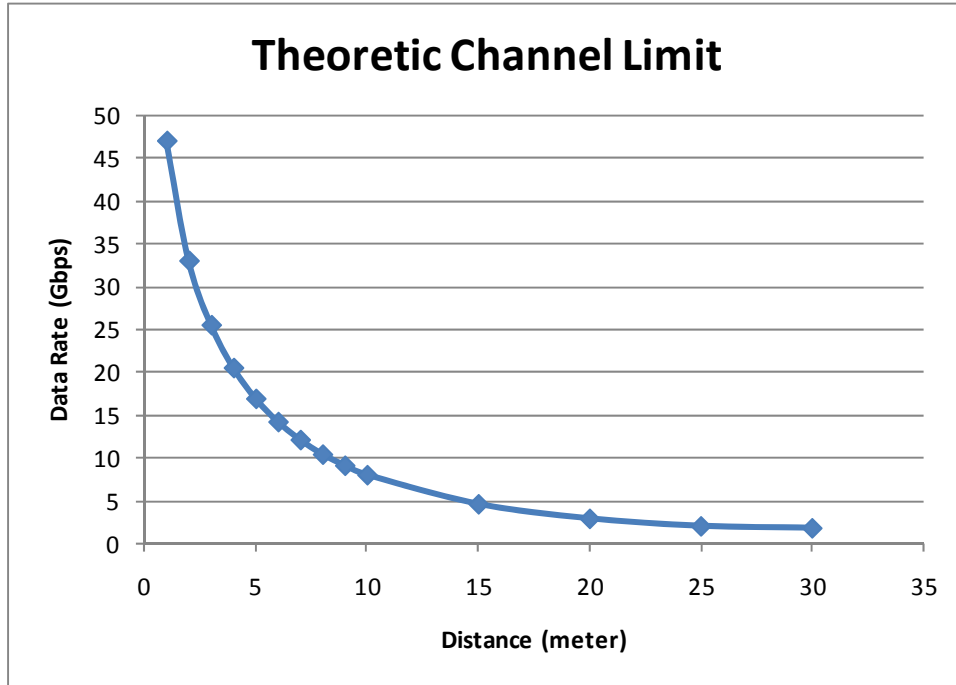


Figure 4.12: Channel capacity of a wireless link built with the octa-core receiver, with a transmitter transmitting at FCC's spectrum mask and isotopic antennas for both RX and TX

Input return loss (3.1~10.6 GHz)		>15 dB ¹
Max. conv. gain	LB	52~61 dB
	MB	52~65 dB
	HB	45~54 dB
VGA range		>40 dB
Noise figure ²	LB	2.6~3.5 dB
	MB	3.4~5 dB
	HB	5~11 dB
Input referred IP3 ³		>-10 dBm
Input referred 1dB compression ³		>-22 dBm
Concurrent cross-band rejection		<-36 dBc
Concurrent core-to-core LO leakage		<-64 dBm
LO spurs		<-26 dBc
Receiver Shannon limit at 5 meters		16 Gbps
Backward compatibility ⁴		group1 MB-OFDM
Power consumption	sleep-mode	1 mA
	WDA and RF buffer	29 mA@1.3 V
	RF balun and distribution	21 mA@1 V
	average unit core ⁵	30 mA@1 V
	max. gain unit core ⁵	55 mA@1 V
	each diff. output	5~25 mA@1 V
	typical total power with N core	29 mA@1.3 V+ (24+40*N) mA@1 V
Technology		130 nm CMOS 1P7M
Die area		1.3 x 2.7 mm ²

All measurement results are based on chip-on-board packaging with 1.0 V supply; otherwise specified:

1. On-wafer measurement result
2. Measured at 1.4V supply with maximum gain setting
3. With minimum gain setting
4. With a 33 MHz LO reference
5. Here, typical gain = maximum gain –15 dB; number includes frequency synthesizer, but excludes output buffers

Table 4.1: Measured performance summary of the octa-core receiver

4.3 Summary

In this chapter, we use concurrency in wireless link to boost communication data rate. As a proof-of-concept, we propose dynamically scalable concurrent communication by dividing the 7.5 GHz bandwidth of the unlicensed 3.1–10.6 GHz spectrum into seven concurrent channels. A CMOS octa-core RF receiver is implemented to validate the idea. Based on the receiver measurement results, a wireless link can be built to achieve a 16 Gbps channel limit at five meter TX-RX distance at 400 mW power consumption.

Chapter 5: Scalable Concurrent Dual-Band Phased Array Receiver

Phased arrays steer beam directions electronically and bring many benefits such as high directivity, interference rejection, signal-to-noise ratio improvement, and fast scanning response [37]–[40]. For this reason, phased arrays have been extensively employed in radar and communication systems in the area of military, space, and radio astronomy since their advent in the 1950s [41] [42]. Recently, substantial attention has also been drawn to civil applications including high-speed point-to-point communications and car radars [40] [43]. However, previous works on phased array IC and system have limited system scalability and diversity. In this chapter, we will propose a scalable concurrent dual-band phased array receiver to relieve these limitations.

This chapter is organized as follows. Section 5.1 briefly reviews phased array systems and the limitations of previous works. The scalable concurrent dual-band phased array receiver architecture will also be proposed in this section. Section 5.2 discusses the difficulty of achieving required tunability using conventional dual-band amplifier topology. In Section 5.3, several tunable concurrent amplifiers will be proposed and compared. Section 5.4 discusses the circuit implementation of major blocks in the tunable dual-band receiver. Section 5.5 presents the experimental results of the receiver test chip and a four-element array system. A chapter summary will be provided in Section 5.6.

5.1. Introduction of Phased Array Receiver

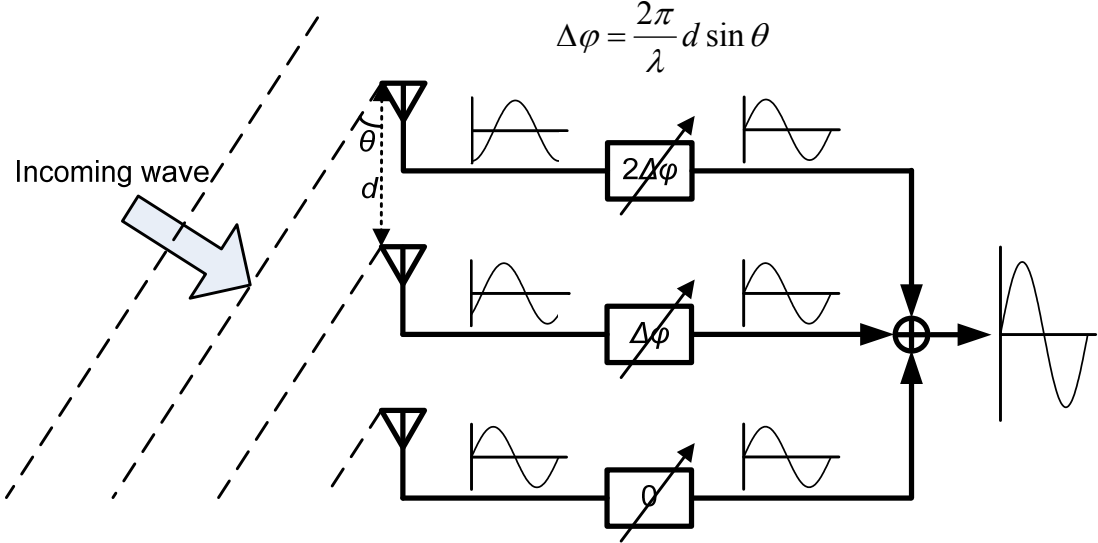


Figure 5.1: Basic phased array receiver configuration

Phased array receivers consist of multiple antenna elements spaced with a certain distance (d) and a following separate phase shifter per element for the electronic beamforming at a given incident angle (θ) in space (Figure 5.1). When a RF wave arrives at the antenna elements, the arrival time of wavefront between two adjacent elements is different by: $\Delta t = \frac{d \cdot \sin \theta}{c}$, where c is the speed of light. In the narrow-band circumstances, the arrival time difference results in a phase delay of the received signal between two adjacent elements, given by: $\Delta\phi = \frac{2\pi d \cdot \sin \theta}{\lambda}$, where λ is the wavelength of the incoming wave. Thus, the following phase shifter adjusts the phase delay in such a way that output signals from each element are all in phase with one another. By summing the signals from each element, a coherent output signal can be obtained with a large array gain. On the other

hand, incoming waves at different incident angles will not be summed coherently. As a result, these signals will be significantly attenuated at the array output.

Since a phased array combines several in-phase signals coherently at the array output, it can achieve an effectively higher gain than a single element receiver. When the signals are combined in the amplitude domain (current or voltage) with a same output load, the array gain is given by:

$$G_{Array} = G_{Single} + 20 \log_{10} N \text{ (dB)} \quad (5.1.1)$$

where G_{Single} is the gain of a single element and N is the number of array elements. Again, undesired signals such as the interference or jammers arriving at other incident angles are inherently rejected according to the established array pattern.

Furthermore, the signal integrity is enhanced at the array output through an effective improvement of the output signal-to-noise ratio (SNR) by a factor of $10 \log_{10} N$ (dB). This is because the noise generated from each element is uncorrelated while the desired signal is combined coherently [22].

Finally, since phased arrays steer beam directions electronically, they are able to receive multiple beams arriving at different incident angles simultaneously. Also, these beams can be steered in a faster and more reliable way than that of a mechanically steered antenna system.

5.1.1. Limitations of Previous Works on Phased Array

The continuing scaling of semiconductor manufacturing not only produces faster transistors, but also allows higher system complexity and integration level. This trend offers an opportunity for dramatic reduction in the cost and the size of phased array systems, in particular in CMOS process. The high yield and repeatability of silicon ICs allows the entire transmitter and/or receiver to be integrated on a single die [22] [23] [24] [25]. This single-chip approach in silicon reduces the overall system cost substantially, compared to the conventional module-based counterpart in compound semiconductors.

The benefits of phased arrays are more noticeable as we increase the number of array elements. Previous works on integrated phased array systems have scalability issues that either the number of phased array elements is limited to the number of array elements that can be implemented inside a single IC, or a large RF signal distribution network will be required in order to combine a very large number of elements, as shown in Figure 5.2. In this figure, several elements are grouped together into a sub-array IC or module, and several sub-arrays are combined by a RF distribution network to present a single output for down-conversion. Therefore, as the number of array elements increases, the cost and complexity of assembling these components into a system will rise dramatically. Furthermore, the design of the low-loss RF distribution network will be challenging with a large number of elements for two reasons: The first reason is that the number of sub-arrays is also increased accordingly, which requires more depth of the signal distribution (or combining) network. The other is that the signal is distributed (or combined) in the RF domain before down-conversion, which gives rise to higher loss than if the distribution (or

combining) were to be performed in the IF or baseband domain. Even more challenges arise when the array must receive multiple beams at the same time. Since each beam requires a separate receiver module and a distribution network for the independent beamforming capability, the associated complexity and cost will be further exacerbated.

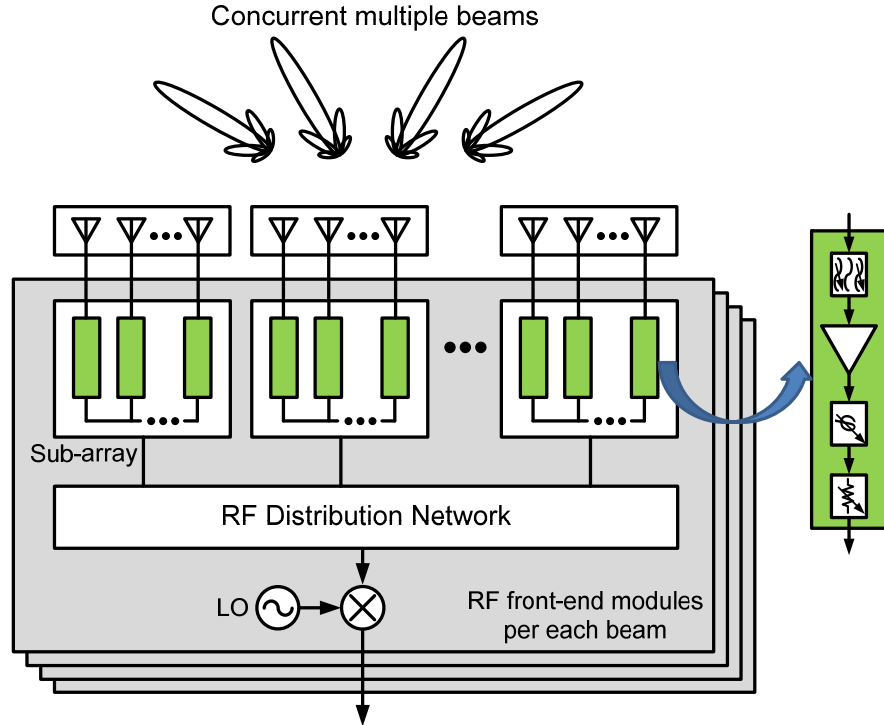


Figure 5.2: A conventional way of building a large-scale phased-array receiver system (in the active array configuration) that supports concurrent multiple beams

Previous works also have limited functionality or diversity. There is a trend in radar and communication systems that the transceivers operate concurrently in multiple modes and multiple bands [26]. Furthermore, many applications require the transceiver to operate in a wide range of RF frequencies [27]. These trends also apply to phased arrays when multiple targets must be tracked at the same time in radar and electronic countermeasure systems or

when multi-point communications are desired at multiple frequencies in a wide bandwidth. The high integration capability of CMOS offers a promising solution to achieve the wideband phased arrays with multiple functionalities. Several wideband phased (or timed) array receivers [28] [29] and transceivers [30] have been reported in silicon. However, none of the previous work has implemented a concurrent multi-band multi-beam phased array receiver operating in a wide range of RF frequencies.

5.1.2. Previous Works on Concurrent Dual-Band Receivers

The fundamental building blocks of the phased array systems are the transceiver elements. The concept of concurrent dual-band operation in radio frequency electronics has been introduced to improve the overall communication throughput and diversity [31]. However, the frequencies of the received RF signals in the previous work are fixed. This limits the application of this architecture to a subset of emerging standards. Concurrent tunability will be studied and introduced in the later sections of this chapter, with IC implementations to prove the concept.

5.1.3. Proposed Large-Scale Phased Array System Architecture

To deal with the scalability issue, we propose an efficient way of building large-scale phased array receiver systems, as shown in Figure 5.3. With a single CMOS chip (a shaded block in Figure 5.3), we integrate all receiver module components on the same die, except for the antenna and front-end LNA. The CMOS receiver includes the tunable concurrent amplifiers (TCAs), down-conversion mixers, phase shifters, frequency synthesizers, and

baseband buffers [32]. This single-chip solution avoids the costly large number of separate component modules and their complicated interconnection for large-scale arrays, which results in a dramatic cost reduction. More importantly, the chip is implemented in CMOS, which will bring another substantial cost reduction compared with its compound-semiconductor counterpart.

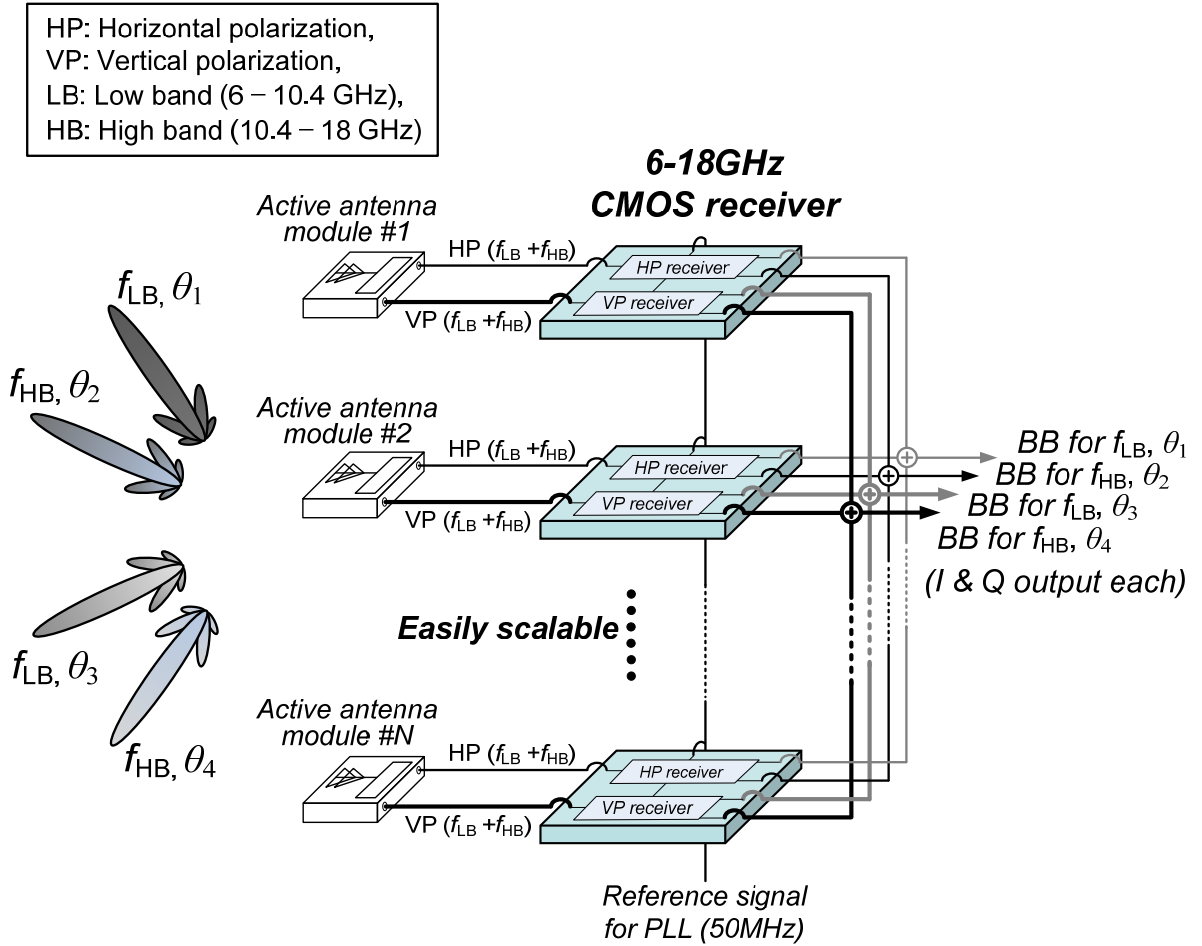


Figure 5.3: A proposed 6–18 GHz phased array receiver system that receives four beams at two frequencies concurrently and is easily scalable toward a very large-scale array

The CMOS receiver has two input ports to receive two different polarization signals fed from an active antenna module, i.e., horizontal polarization (HP) and vertical polarization (VP), respectively. On the other hand, each input port is able to receive a dual-band signal containing two different frequencies concurrently, one in the low band (LB) from 6 to 10.4 GHz and the other in the high band (HB) from 10.4 to 18 GHz. The dual-band signal is then split into two separate signals on-chip, one for each band. Subsequently, each signal is down-converted with the independent phase-shifting operation to provide separate beamforming. Therefore, the proposed array system can receive and steer four different beams at two different frequencies concurrently.

The baseband outputs from each array element are combined off-chip in the current domain, providing the back-end processors with one combined baseband signal per beam. Since the signal combining is performed at the baseband rather than at the RF frequency, it alleviates the difficulty in designing a low-loss combining network for a large-scale array.

It is also noteworthy that the 50 MHz LO references signal is the only signal which needs to be distributed among the elements other than DC supplies. Due to its low frequency, the reference can be simply distributed without adding complexity. It also makes the proposed array architecture easily scalable.

The LO signals generated by the on-chip frequency synthesizers may have relatively higher phase noise than those provided by off-chip low-noise sources. However, when combining N elements (or N chips) in the array, the phase noise originating from the on-chip components of each element is uncorrelated with one another and thus adds up in power. On the other hand, the carrier signal is combined in amplitude in the current

domain. Therefore, the phase noise performance is improved by a factor of $10 \log_{10} N$ (dB) at the array output. This improvement also makes the single-chip solution, including on-chip frequency synthesizers, suitable for large-scale phased arrays without degrading the array performance.

In the complete array system, a separate active antenna module, consisting of a broadband antenna and a GaN LNA, will be employed in front of the CMOS receiver.

5.2. Tunability of Concurrent Dual-Band Amplifiers

One of the major challenges in implementing the proposed phased array systems introduced in Section 5.1.3 is the implementation of the dual-band TCA. One of the possible solutions is to make a conventional concurrent dual-band amplifier tunable. A concurrent dual-band amplifier has a dual-resonant input matching network, and a dual-resonant output network, as shown in Figure 5.4. For concurrent operations, resonant frequencies of these two networks need to be matched.

To make the dual-band amplifier in Figure 5.4 tunable, variable capacitors (C_g , C_{gs} , C_1 , and C_2) need to be implemented, which have a limited tuning range (e.g., MOS varactors have a typical tuning range of three). Under this constraint, a tunable dual-resonant input matching network and output network can be designed to cover any frequency between 6 and 18 GHz by either of the pass-bands of the dual-resonant networks.

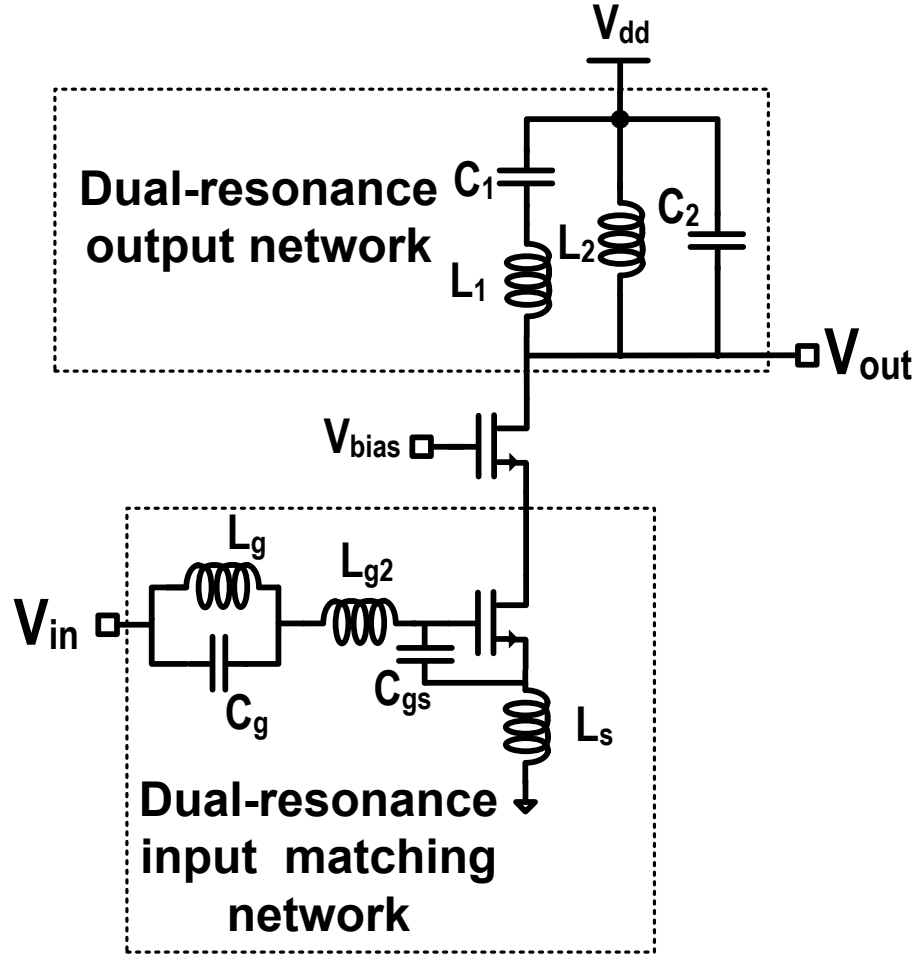


Figure 5.4: Schematics of a concurrent dual-band amplifier

If we look at the output load resonant network as an example, we will find that the output network has two resonant peaks at $\omega_{L,load}$, and $\omega_{H,load}$, respectively, where

$$\omega_{L,load} = \sqrt{\frac{(L_1 C_1 + L_2 C_1 + L_2 C_2) - \sqrt{(L_1 C_1 + L_2 C_1 + L_2 C_2)^2 - 4 L_1 L_2 C_1 C_2}}{2 L_1 L_2 C_1 C_2}} \quad (5.2.1)$$

and

$$\omega_{H,load} = \sqrt{\frac{(L_1 C_1 + L_2 C_1 + L_2 C_2) + \sqrt{(L_1 C_1 + L_2 C_1 + L_2 C_2)^2 - 4 L_1 L_2 C_1 C_2}}{2 L_1 L_2 C_1 C_2}}. \quad (5.2.2)$$

Here, $\omega_{H,load} > \omega_{L,load}$, and they are the two passband frequencies of the output concurrent load networks.

Both $\omega_{H,load}$ and $\omega_{L,load}$ are functions of C_1 and C_2 which have limited tuning range, said $C_1 \in \{C_{1,min}, C_{1,max}\}$ and $C_2 \in \{C_{2,min}, C_{2,max}\}$. Rewrite $\omega_{L,load}$ as $\omega_{L,load}(C_1, C_2)$ and $\omega_{H,load}$ as $\omega_{H,load}(C_1, C_2)$. We would like:

$$\begin{aligned} \omega_{L,load}(C_{1,max}, C_{2,max}) &\leq 2\pi \cdot 6 \text{ GHz} \\ \omega_{H,load}(C_{1,min}, C_{2,min}) &\geq 2\pi \cdot 18 \text{ GHz} \\ \omega_{L,load}(C_{1,min}, C_{2,min}) &> \omega_{H,load}(C_{1,max}, C_{2,max}) \end{aligned} \quad (5.2.3)$$

to be satisfied, so that any frequency between 6 to 18 GHz can be covered by either of the pass bands. Solving the inequalities of Equation (5.2.3) gives the design values for the passive elements inside the load network. A similar process and solution set can be derived for the input-matching network to get its pass-band frequencies: $\omega_{L,input}(C_g, C_{gs})$ and $\omega_{H,input}(C_g, C_{gs})$, as well as design value for (C_g, C_{gs}) .

One way to illustrate the tunability of a concurrent dual-band network is to plot all combinations of $[\omega_{L,load}(C_1, C_2), \omega_{H,load}(C_1, C_2)]$ on a 2-D coordinate system by sweeping C_1 between $C_{1,min}$ and $C_{1,max}$, and C_2 between $C_{2,min}$ and $C_{2,max}$. Similarly, we plot $[\omega_{L,input}(C_g, C_{gs}), \omega_{H,input}(C_g, C_{gs})]$ on another 2-D coordinate system by sweeping C_g between $C_{g,min}$ and $C_{g,max}$, and C_{gs} between $C_{gs,min}$ and $C_{gs,max}$. The gray area

enclosed by thick black lines in Figure 5.5 represents the achievable region of the concurrent dual-band operation.

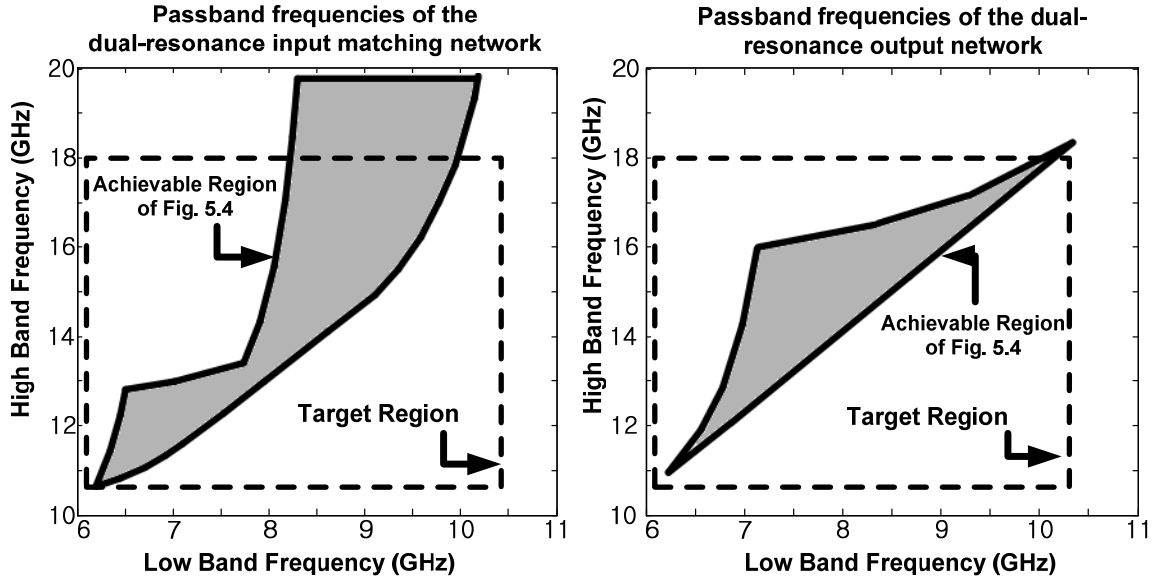


Figure 5.5: Achievable frequency region of tunable dual-band operation of the amplifiers in Figure 5.4 with limited capacitor tuning range, and all frequencies between 6–18 GHz covered by either band

There are three major problems in making the architecture in Figure 5.4 tunable: Firstly, the high-band and the low-band frequencies cannot be independently controlled. Secondly, the achievable region of operation is only a small portion of the desired operation region, which is the rectangle enclosed by dotted lines. Thirdly, matching the resonant frequencies of the two networks is difficult. Furthermore, it is not clear how or if it is possible to achieve this with higher-order networks due to resonant couplings in the system. These conclusions also suggest that to achieve an independent tunability between the high band and the low band, isolation between the high-band and the low-band resonators is

necessary. Based on these preliminary studies, in the next section we will explore several possible topologies of TCA which comprise two isolated resonators.

5.3. Tunable Concurrent Amplifier (TCA)

The TCA needs to provide good broadband input matching and good isolation between its two outputs. In this section, we will compare the dynamic range performance of four different TCA topologies, i.e., the common-gate common-gate topology, the common-gate common-source topology, the resistor-terminated topology, and the active-termination topology. All of these four topologies satisfy our basic design requirements. The inductive degeneration-based architecture [33] is not compared here, because the effective transconductance of an inductively degenerated amplifying core presents a non-flat inherent frequency response.

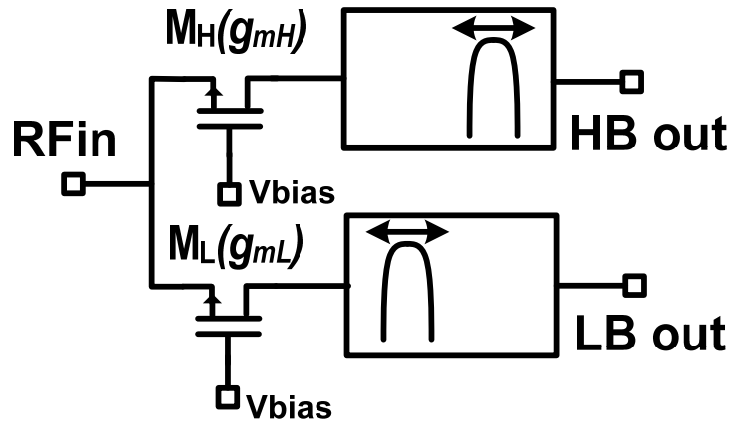


Figure 5.6: Schematics of a common-gate common-gate TCA

5.3.1. Common-Gate Common-Gate (CG-CG) Topology

The CG-CG topology matches its input impedance by the two parallel CG amplifiers. Each of the CG has an $1/g_m$ input impedance, where g_m is its transconductance. The effective input impedance of a CG-CG TCA is:

$$Z_{in} = \frac{1}{g_{mL} + g_{mH}}. \quad (5.3.1)$$

A CG amplifier has very good isolation between its source and drain node, so the CG-CG TCA has a good isolation between its two outputs. When input is matched ($Z_{in} = Z_0$), the noise figure of this TCA is:

$$\begin{aligned} NF_{HB} &\approx 1 + Z_0 g_{mL} \gamma + \frac{g_{mH} \gamma}{g_{mL} + g_{mH}} \cdot \left\{ 1 + \frac{2g_{mL}}{g_{mH}} \right\}^2 \\ NF_{LB} &\approx 1 + Z_0 g_{mH} \gamma + \frac{g_{mL} \gamma}{g_{mL} + g_{mH}} \cdot \left\{ 1 + \frac{2g_{mH}}{g_{mL}} \right\}^2. \end{aligned} \quad (5.3.2)$$

NF_{HB} and NF_{LB} are the HB and LB noise figures, respectively. When $g_{mH} \approx g_{mL}$, the above equation can be simplified into:

$$NF \approx 1 + 5\gamma. \quad (5.3.3)$$

The noise figure of CG-CG stage is bad for two reasons: Firstly, to match the input impedance using two CG amplifiers, both of the CG amplifiers will have a smaller transconductance which dramatically decreases the gain of the amplifiers; noise generated from other sources will be more dominating than the source impedance. Secondly, thermal noise from the HB CG transistor will leak into the LB path, and vice versa.

The P1dB of this topology is:

$$P_{1dB} \approx \frac{2.9|g_m| \cdot \Omega^{-1}}{|g_{m3}|} \text{ (mW)}. \quad (5.3.4)$$

g_{m3} is the third-order transconductance of the transistor. g_m/g_{m3} has a unit of Volt^2 , so we need to multiply it by $\text{Ohm}^{-1} = \text{Volt}^{-1} \cdot \text{Ampere}$ to make P_{1dB} 's unit correct.

The dynamic range (DR) performance of the CG-CG TCA will then be:

$$DR \approx \frac{2.9 \times 10^{-3} \cdot |g_m| \cdot \Omega^{-1}}{|g_{m3}| \cdot kT \cdot (1+5\gamma) \cdot BW}. \quad (5.3.5)$$

BW is the bandwidth of the front-end filter. k is *Boltzmann's* constant, and T is the temperature of the chip. The DR can be improved by reducing the noise figure performance of the CG-CG TCA.

5.3.2. Common-Gate Common-Source (CG-CS) Topology

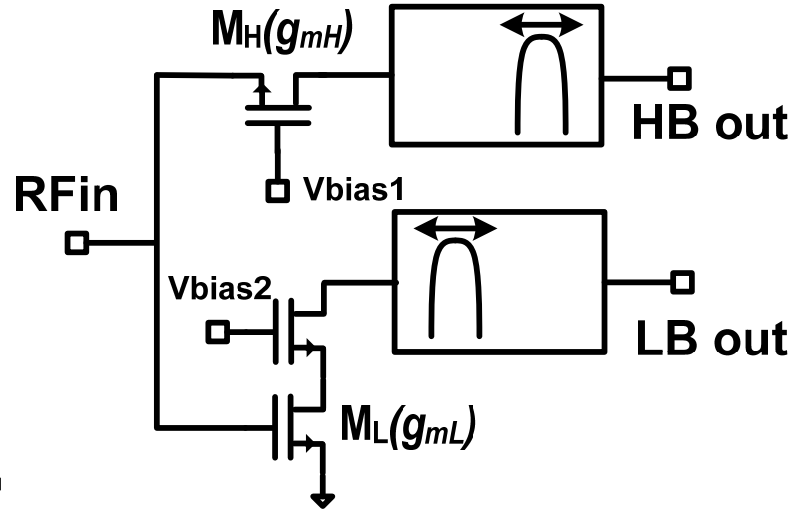


Figure 5.7: Schematics of a common-gate common-source TCA

The CG-CS topology matches its input impedance by a CG amplifier. The input impedance transforms the input RF power into a gate voltage, which is amplified by both the CG transistor (shown as the HB path in Figure 5.7), and the CS transistor (shown as the LB path in Figure 5.7). Both the CG and the cascode amplifier have very good isolation from input to output, so isolation between the HB and LB outputs can be achieved. When the input is matched ($Z_{in} = Z_0 = \frac{1}{g_{mH}}$), the noise figure of this CG-CS TCA is:

$$\begin{aligned} NF_{HB} &\approx 1 + \gamma \\ NF_{LB} &\approx 1 + \gamma + \frac{4\gamma}{g_{mL}Z_0}. \end{aligned} \quad (5.3.6)$$

NF_{LB} is greater than NF_{HB} , because drain noise from the CG transistor leaks to the LB signal path. However both noise figures are better than that in the CG-CG topology. Input matching is achieved by a single CG stage, and the effective transconductance of both HB and LB paths can be increased. In addition, g_{mL} is decoupled from the input matching, so we can choose a large g_{mL} to reduce the last term in NF_{LB} equation. The P_{1dB} of the CG-CS TCA is comparable to the CG-CG TCA, and the dynamic range of the CG-CS TCA is:

$$\begin{aligned} DR_{HB} &\approx \frac{2.9 \times 10^{-3} \cdot |g_{mH}| \cdot \Omega^{-1}}{|g_{mH3}| \cdot kT \cdot (1 + \gamma) \cdot BW} \\ DR_{LB} &\approx \frac{2.9 \times 10^{-3} \cdot \Omega^{-1}}{\left| \frac{2g_{mL3}}{g_{mL}} - \frac{g_{mH3}}{g_{mH}} \right| \cdot kT \cdot \left(1 + \gamma + \frac{4\gamma}{g_{mL}Z_0} \right) \cdot BW}. \end{aligned} \quad (5.3.7)$$

g_{mH3} and g_{mL3} are the third-order transconductance of the HB and LB transistors, respectively.

5.3.3. Resistor Termination Topology

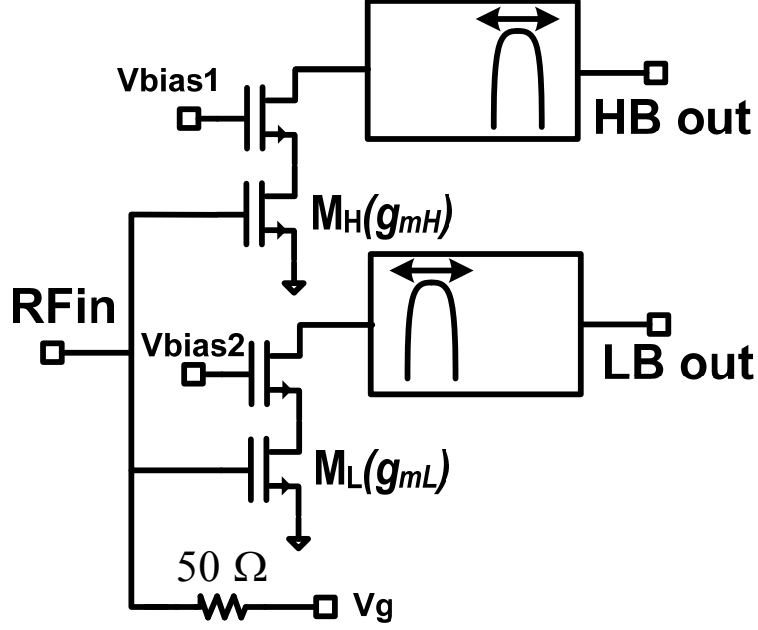


Figure 5.8: Schematics of a resistor-terminated TCA

The resistor-terminated TCA matches its input impedance with a $50\ \Omega$ resistor, as shown in Figure 5.8. This $50\ \Omega$ resistor transforms the input RF power into a gate voltage, which is amplified by the cascode amplifiers for both the HB and LB signal paths. Since the cascode amplifiers have good isolation between the inputs and outputs, isolation between the HB and LB outputs can be achieved. The noise figure of this resistor-terminated TCA is:

$$NF_{HB} \approx 2 + \frac{4kT}{g_{mH}Z_0}$$

$$NF_{LB} \approx 2 + \frac{4kT}{g_{mL}Z_0}.$$
(5.3.8)

Since the g_{mL} and g_{mH} are decoupled from input matching, we can choose a large value to reduce the second terms in Equation (5.3.8). The P_{1dB} of the resistor-terminated TCA is similar to the CG-CG TCA, so the dynamic range of the CG-CS TCA is:

$$DR_{HB} \approx \frac{1.5 \times 10^{-3} \cdot |g_{mH}| \cdot \Omega^{-1}}{|g_{mH3}| \cdot kT \cdot \left(2 + \frac{4kT}{g_{mH}Z_0}\right) \cdot BW} \quad (5.3.9)$$

$$DR_{LB} \approx \frac{1.5 \times 10^{-3} \cdot |g_{mL}| \cdot \Omega^{-1}}{|g_{mL3}| \cdot kT \cdot \left(2 + \frac{4kT}{g_{mL}Z_0}\right) \cdot BW}$$

5.3.4. Active Termination Topology

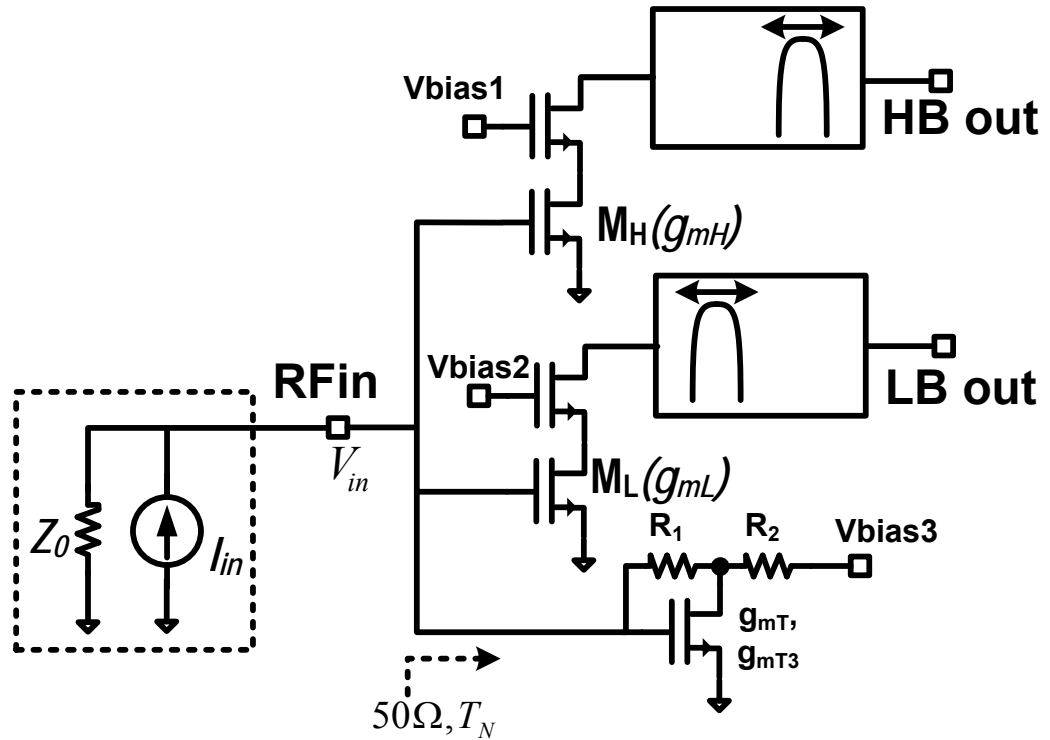


Figure 5.9: Schematics of an active-termination TCA

The noise figure of a resistor-terminated TCA is lower bounded by the thermal noise generated from the $50\ \Omega$ termination resistor. A possible way to reduce this lower bound is by using the active termination [34], which effectively makes a lower-than-room-temperature $50\ \Omega$ small-signal resistance. The detailed schematic of this TCA is shown in Figure 5.9.

The noise temperature T_N of an active termination at match condition ($Z_0 = \frac{R_1 + R_2}{1 + g_{mT}R_2}$, $Z_0 = 50\ \Omega$) can be found to be:

$$T_N = \frac{1}{1 + g_{mT}R_2} + \frac{g_{mT}R_2^2\gamma}{Z_0(1 + g_{mT}R_2)^2}. \quad (5.3.10)$$

R_1 is a function of g_{mT} and R_2 when input impedance is made to be Z_0 . Observe equation, since all parameters are positive $T_N > 0$. Also, if we take the limit: $\lim_{g_{mT} \rightarrow +\infty} T_N = 0$, we know that T_N has a minimum at $g_{mT} \rightarrow +\infty$. And $T_N = 1$, for $g_{mT} = 0$. So T_N will have a global maximum between $g_{mT} \in [0, +\infty)$, for finite R_2 .

For $R_2 = 0$, $T_N = 1$, and $\lim_{R_2 \rightarrow +\infty} T_N = \frac{\gamma}{Z_0 g_{mT}}$. There exists a global minimum in $R_2 \in [0, +\infty)$, if $Z_0 g_{mT} < \gamma$, and a global maximum if $Z_0 g_{mT} > \gamma$ for a finite g_{mT} . Take the derivative of T_N by R_2 :

$$\frac{\partial T_N}{\partial R_2} = -\frac{g_{mT}}{(1 + g_{mT}R_2)^2} + \frac{2g_{mT}R_2\gamma}{Z_0(1 + g_{mT}R_2)^2} - \frac{2g_{mT}^2R_2^2\gamma}{Z_0(1 + g_{mT}R_2)^3}. \quad (5.3.11)$$

Solving $\frac{\partial T_N}{\partial R_2} = 0$, we will get: $R_2 = \frac{Z_0}{2\gamma - g_{mT}Z_0}$. Since R_2 needs to be positive, this means

$g_{mT}Z_0 < 2\gamma$. Apply R_2 into Equation (5.3.10) and we will get:

$$T_N = 1 - \frac{g_{mT}Z_0}{4\gamma} < 1. \quad (5.3.12)$$

Since $T_N < 1$, it cannot be a global maximum. To verify that T_N is a global minimum at

$R_2 = \frac{Z_0}{2\gamma - g_{mT}Z_0}$, we need to the derivative of Equation (5.3.11):

$$\begin{aligned} \frac{\partial^2 T_N}{\partial R_2^2} = & \frac{2g_{mT}^2}{(1+g_{mT}R_2)^3} + \frac{2g_{mT}\gamma}{Z_0(1+g_{mT}R_2)^2} - \frac{4g_{mT}^2R_2\gamma}{Z_0(1+g_{mT}R_2)^3} - \frac{4g_{mT}^2R_2\gamma}{Z_0(1+g_{mT}R_2)^3} + \\ & \frac{6g_{mT}^3R_2^2\gamma}{Z_0(1+g_{mT}R_2)^4}. \end{aligned} \quad (5.3.13)$$

And applying $R_2 = \frac{Z_0}{2\gamma - g_{mT}Z_0}$ to the above equation, we will get:

$$\frac{\partial^2 T_N}{\partial R_2^2} = \frac{g_{mT}}{2Z_0\gamma} \cdot (2\gamma - g_{mT}Z_0)^2 \cdot \left(1 - \frac{Z_0g_{mT}}{2\gamma}\right)^2 > 0. \quad (5.3.14)$$

So, we prove that $T_{N,min} = 1 - \frac{g_{mT}Z_0}{4\gamma}$ at $R_2 = \frac{Z_0}{2\gamma - g_{mT}Z_0}$ for $g_{mT} < \frac{2\gamma}{Z_0}$. For $g_{mT} \geq \frac{2\gamma}{Z_0}$,

$T_{N,min} = \frac{\gamma}{Z_0g_{mT}}$ at $R_2 \rightarrow +\infty$, which can approach zero when $g_{mT} \rightarrow +\infty$.

The noise figure of both signal paths are now:

$$\begin{aligned} NF_{HB} &\approx 1 + \frac{T_N}{T_0} + \frac{4kT}{g_{mH}Z_0} \\ NF_{LB} &\approx 1 + \frac{T_N}{T_0} + \frac{4kT}{g_{mL}Z_0}. \end{aligned} \quad (5.3.15)$$

We will find that the noise figure of the active-termination TCA is better than any of the previously discussed topologies. When we drive the active-termination TCA with a power

source (shown as a source impedance Z_0 shunt with a constant current source I_{in}), we can calculate the gate voltage V_{in} with respect to the input current I_{in} as:

$$V_{in} \approx \frac{Z_0}{2} I_{in} + \frac{|g_{mT3}| R_2 Z_0^3}{16(1+g_{mT} R_2)} I_{in}^3. \quad (5.3.16)$$

g_{mT3} is the third-order transconductance of the active termination transistor. The TCA output current can be calculated to be:

$$I_{out} = \frac{Z_0 g_m}{2} I_{in} + \left(\frac{g_m |g_{mT3}| R_2}{16(1+g_{mT} R_2)} - \frac{|g_{m3}|}{8} \right) Z_0^3 I_{in}^3. \quad (5.3.17)$$

The P_{1dB} of the TCA will now become:

$$P_{1dB} = \frac{2.9 \cdot \Omega^{-1}}{\left(\frac{|g_{mT3}| R_2}{(1+g_{mT} R_2)} - \frac{2|g_{m3}|}{g_m} \right)} (mW). \quad (5.3.18)$$

Comparing Equation (5.3.18) with Equation (5.3.4), we find that active-termination TCA has a better P_{1dB} , because the third-order nonlinearity from the active-termination partially cancels the third-order nonlinearity from the main amplifying transistors. The DR of the active-termination TCA is hence the largest among all discussed topologies. The DR of the active-termination TCA is:

$$DR \approx \frac{2.9 \times 10^{-3} \cdot \Omega^{-1}}{\left(\frac{|g_{mT3}| R_2}{(1+g_{mT} R_2)} - \frac{2|g_{m3}|}{g_m} \right) \cdot kT \cdot \left(1 + \frac{T_N}{T_0} + \frac{4kT}{g_{mH} Z_0} \right) \cdot BW}. \quad (5.3.19)$$

Since the active-termination TCA has the largest DR, we choose this topology in our final IC implementation.

5.4 A 6–18 GHz Concurrent Tunable Dual-Band Phased Array Receiver

In this section, the architecture and frequency plan of the CMOS concurrent phased array receiver element is discussed in detail. It should be noted that a single receiver chip operates as one receiver element in the array system, as shown in Figure 5.3.

5.4.1 Block Diagrams

A block diagram of the receiver architecture is presented in Figure 5.10. Since it is a concurrent dual-band receiver, the incoming RF signal contains two frequencies at LB and HB, respectively, and feeds a front-end tunable concurrent amplifier (TCA). The TCA amplifies, filters, and finally splits the RF signal into two separate outputs: one at LB and the other at HB. Each of the two signals goes through separate double down-conversion by subsequent RF and IF mixers. The IF mixers generate the I and Q components of the baseband signal for digital demodulation capability. The baseband VGAs adjust the baseband amplitude and drive the output load differentially.

There are two sets of RF input (HP RF input and VP RF input in Figure 5.10) which are down-converted by two same sets of the RF signal-path circuitry, respectively. Therefore, the receiver presents a total of eight differential baseband outputs, one for each combination of two different polarizations (HP and VP), two different frequency bands (LB and HB), and I and Q.

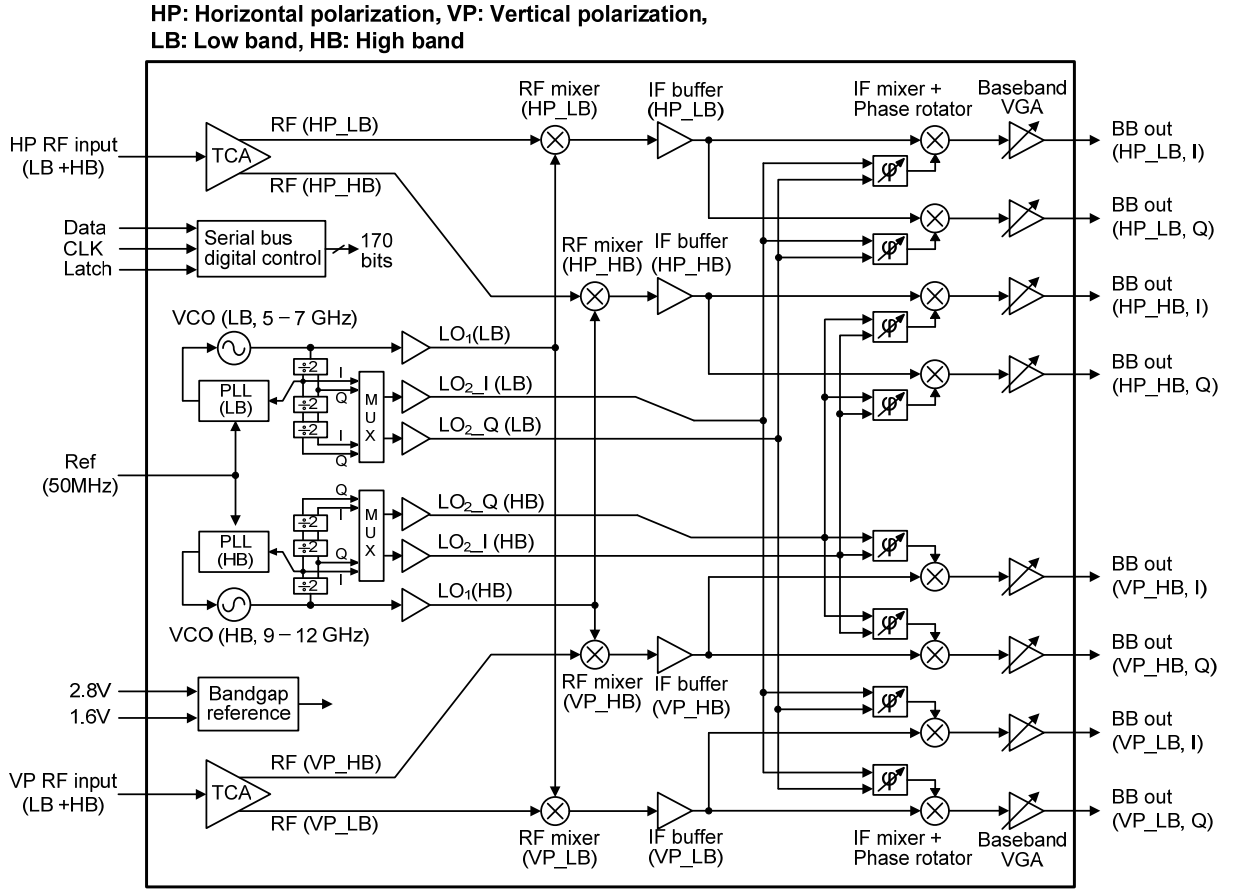


Figure 5.10: Architecture of the tunable concurrent dual-band quad-beam phased array receiver in CMOS

The receiver includes two on-chip programmable frequency synthesizers in order to support the separate down-conversion of the LB and HB signals, respectively. The frequency synthesizers generate the first LO (LO_1) signal between 5–7 GHz for LB and between 9–12 GHz for HB with a frequency step of 200 MHz. The LO_1 signal drives the RF mixers for two polarizations. The second LO (LO_2) signal, driving the phase rotators and IF mixers, is generated by three static divide-by-2 dividers and a 2:1 multiplexer. According to the receiver frequency scheme shown in Figure 5.11, the LO_2

frequency is selected as either one half or one eighth of the LO_1 frequency by the multiplexer. The LO_2 signal carries the I and Q components separately to feed the phase rotators in quadrature. A 50 MHz reference signal for the PLLs is generated by an off-chip crystal oscillator.

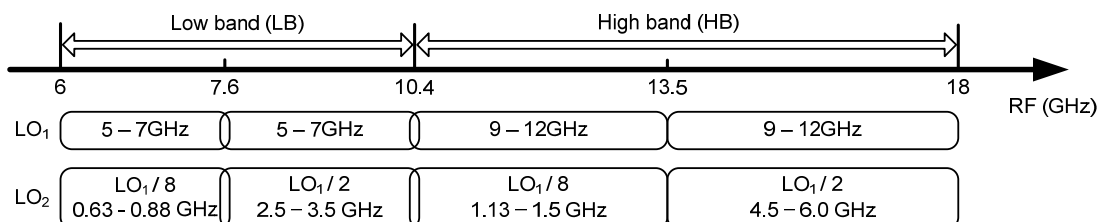


Figure 5.11: Frequency scheme

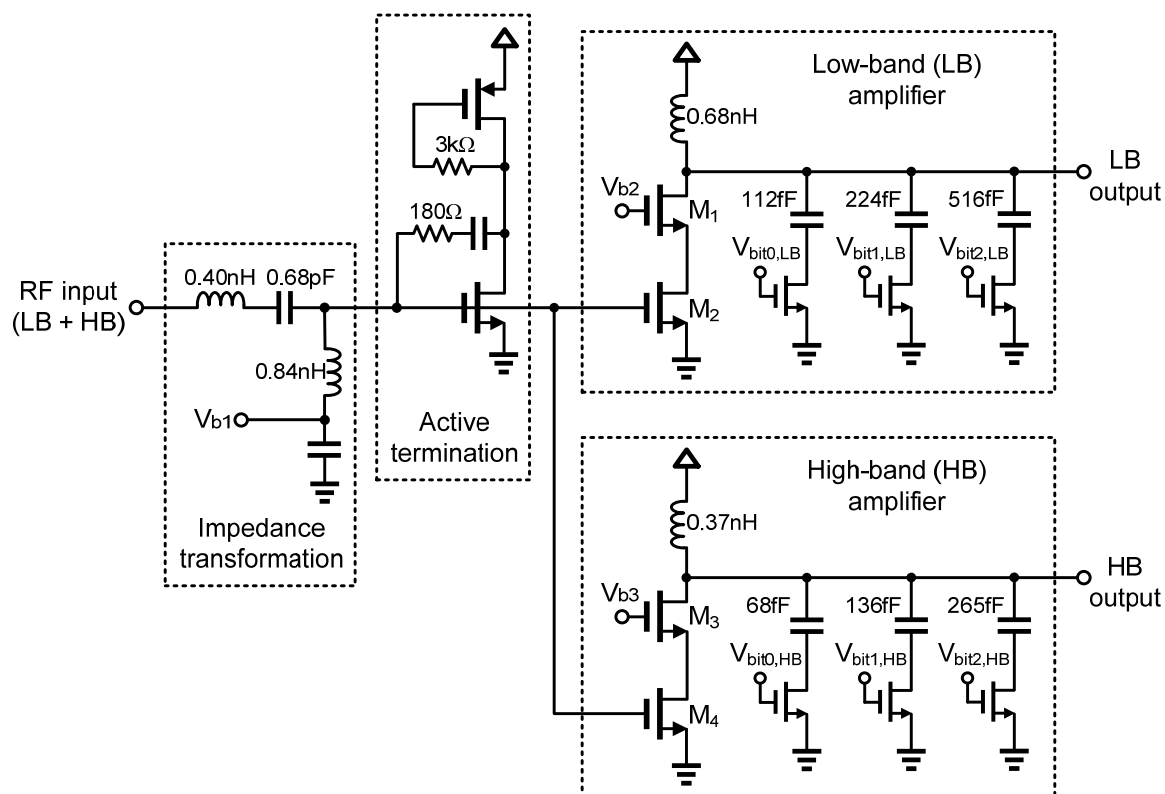


Figure 5.12: Schematic of the TCA with a single input and a dual output

The LO phase-shifting architecture is adopted in this phased array receiver in order to circumvent the challenge of designing high-resolution wideband phase shifters in the RF signal path [35]. The phase shifting is performed in the LO_2 signal by a 10 bit digital phase rotator. Each IF mixer is driven by a separate phase rotator to maximize the flexibility of the receiver. This not only provides independent beamforming capability to the signals of different bands and polarizations, but also helps to minimize the I and Q mismatch of the quadrature baseband outputs.

The receiver includes an on-chip digital serial-bus control unit that programs 170 bits to configure the dual RF frequencies, LO frequencies, phase-shifting angles, baseband gains, and other functionalities of the receiver. Bias voltages are generated by on-chip bandgap reference circuitry.

5.4.2 TCA

Based on the discussion in Section 5.3, the final schematic of the TCA is shown in Figure 5.12. An input-matching (impedance transformation) network is added in front of the active-termination TCA to relieve the input-matching degradation due to the parasitic capacitance seen at the input node of the TCA. The active termination not only provides the required resistive impedance, but also produces less thermal noise compared to a $50\ \Omega$ resistor. The input RF signal power is converted to a signal voltage by the active termination, and this voltage is amplified by two cascode amplification stages. The cascode amplifiers not only enhance the isolation between the two output signals, but also minimize the crosstalk of noise produced by the active blocks. The RF signals at two

frequencies are then selectively amplified by two separate cascode amplifiers (M_1 – M_2 , M_3 – M_4) that have tunable LC output loads. A 3 bit switched capacitor bank at each output load is tuned to cover the entire LB and HB frequencies. This allows for the digital tuning of the amplifier so that it can provide the maximum gain at the desired frequency and attenuate out-of-band signals prior to the first down-conversion.

5.4.3 RF and IF Mixers

Four different mixer designs are presented in the receiver: RF and IF mixers, each for LB and HB. The current-commutating double-balanced topology is adopted for all the mixers in order to minimize the LO-to-IF feedthrough. Figure. 5.13 shows the schematic of the RF mixer and IF buffer for LB. A shunt-peaking inductor (3.3 nH) is used to extend the IF 3 dB bandwidth up to over 3.5 GHz. Since the TCA provides a single-ended RF signal to the differential RF mixers, one RF input terminal is terminated to a bias voltage by a 2 k Ω resistor and a bypass capacitor.

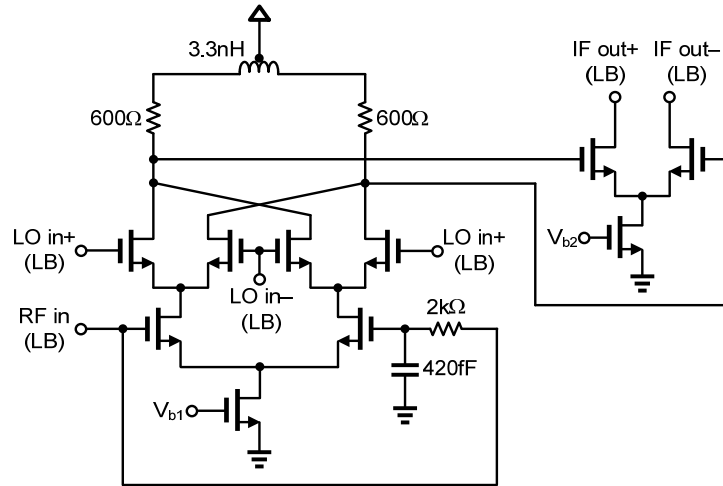


Figure 5.13: Schematic of the RF mixer and IF buffer for LB

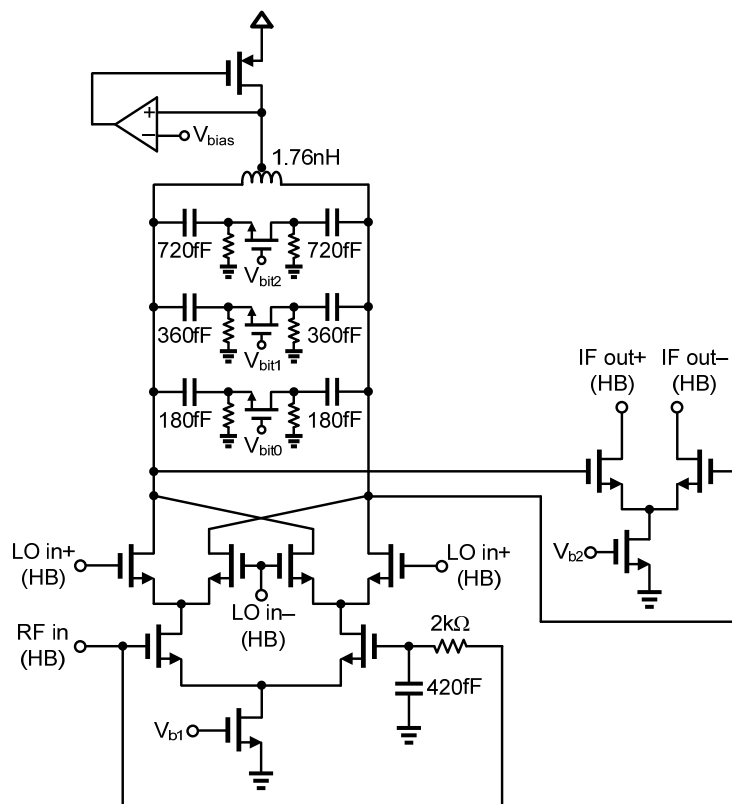


Figure 5.14: Schematic of the RF mixer for HB

The HB RF mixer employs a tunable LC load with a 3-bit switched capacitor bank at the IF output, as shown in Figure 5.14. The resonant frequency of the LC load is tuned in such a way that the conversion gain is maximized at the desired IF frequency. The common-mode feedback circuitry ensures a given bias voltage (V_{bias}) set for the subsequent buffer block.

The schematics of the IF mixers for LB and HB are similar to that of the LB RF mixer. However, the IF mixers employ no shunt-peaking inductors and are degenerated by source resistors to improve linearity of the baseband signal.

5.4.4 Baseband Buffers

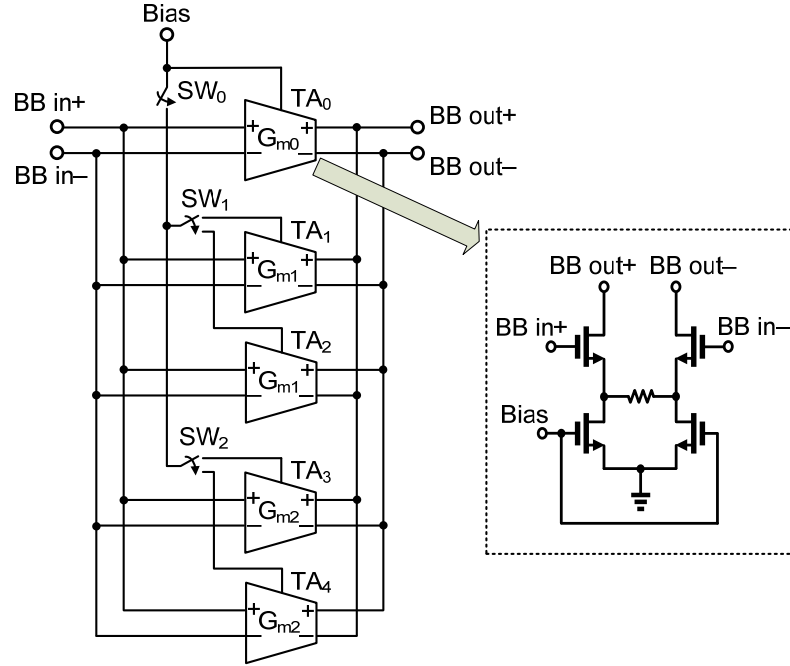


Figure 5.15: Schematic of baseband VGA

The VGA combines five transconductance amplifiers in the current domain with digitally switched bias voltages (Figure. 5.15). TA_1 and TA_2 , TA_3 and TA_4 are identical pairs that constitute current-commutating cells by digital switches (SW_1 and SW_2). Each transconductance amplifier has a differential common-source topology with resistive degeneration. Since the output port is configured with open drains, the output signals from each array element can be easily combined in the current domain using a passive network which imposes little additional impact on the nonlinearity performance. The open-drain output requires an external DC supply of 1.5 V. The VGA achieves a nominal gain of 7 dB, with a 11 dB gain variation in five steps when driving a 100 Ω differential output load.

5.4.5 Whole Receiver Chip

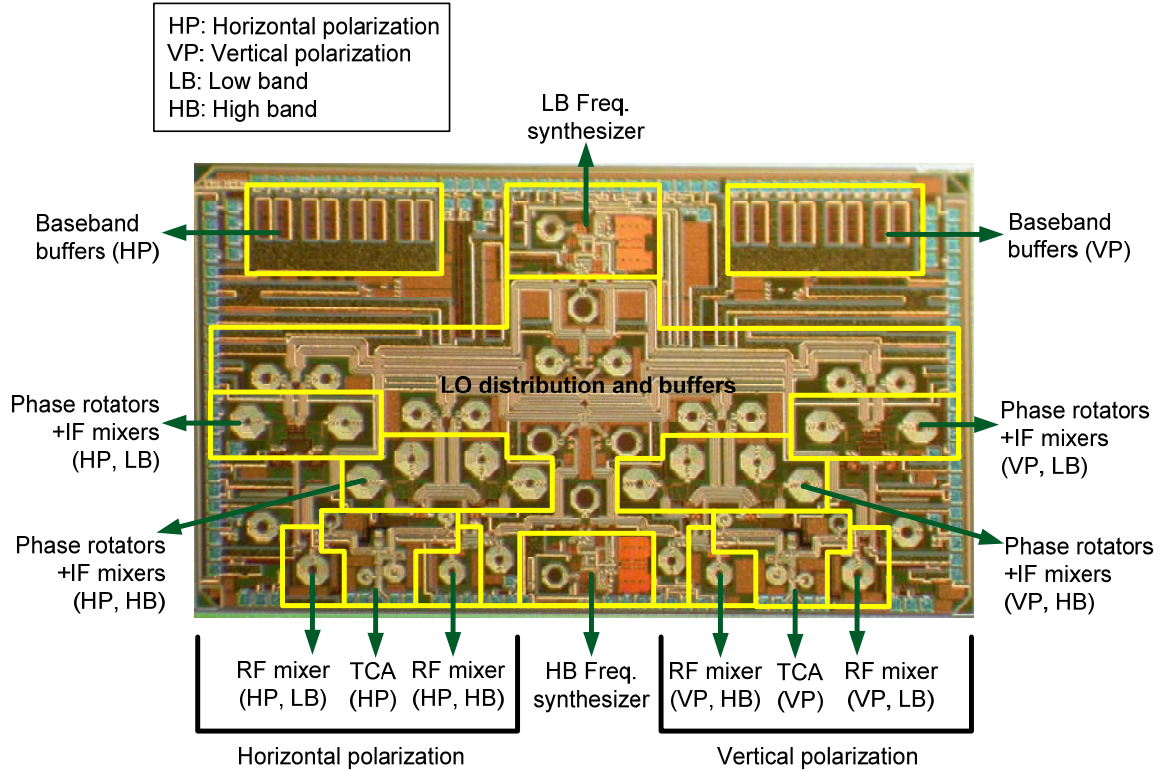


Figure 5.16: Chip micrograph

The phased-array receiver element is implemented in a 130 nm CMOS process. It provides eight metal layers, including two top thick metal layers of 4 μm aluminum and 3- μm copper. Figure 5.16 shows a die micrograph of the implemented chip that occupies an area of $3.0 \times 5.2 \text{ mm}^2$.

5.5 Experimental Results

5.5.1 Receiver Test Circuits

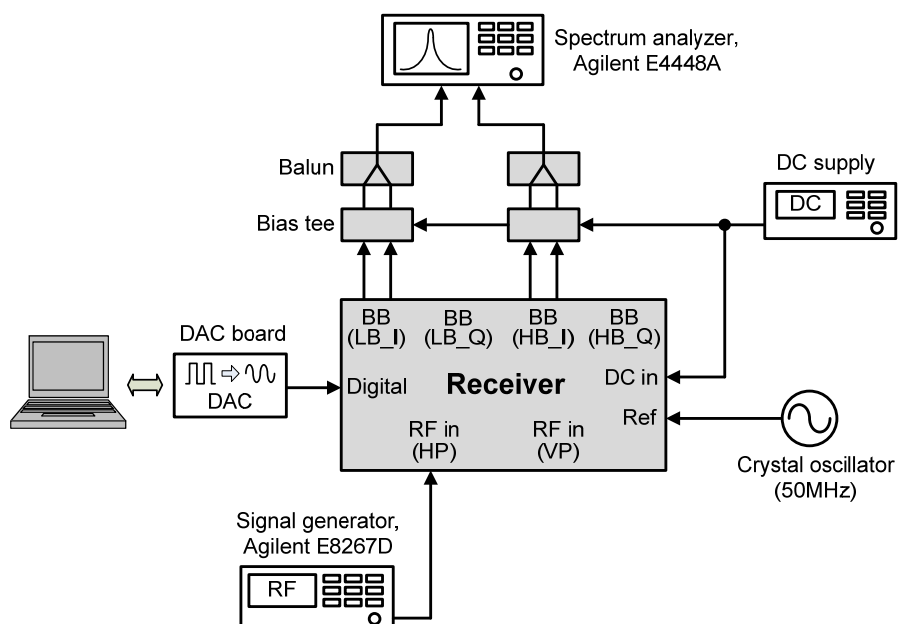


Figure 5.17: Block diagram of the receiver measurement setup

For the measurement of the receiver element, a printed circuit board (PCB) is designed on a Duroid substrate of 0.254 mm thickness. The PCB provides the traces for the DC supplies, reference signal, digital signals, and differential baseband outputs. All signal inputs and outputs are fed with SMA connectors. The PCB is attached on a gold-plated brass board. Then, through a pre-cut aperture of the PCB, the chip is mounted directly on the brass board using silver epoxy in order to provide good substrate grounding and heat sink. The chip pads are wire-bonded to the PCB traces, except that the ground pads are wire-bonded directly to the brass board.

A block diagram of the measurement setup is shown in Figure 5.17. The RF input signal is fed by a coplanar GSG probe to minimize the feed loss. Off-chip baluns convert the differential baseband output to a single-ended one for measurement purposes. There are three different DC supplies applied to the chip; 1.6 V and 2.7 V for the RF and LO circuitry, and 1.5 V for the baseband buffers. A temperature-compensated crystal oscillator with phase noise of -155 dBc/Hz at 1 kHz offset provides a 50 MHz reference signal for the on-chip PLLs. Digital codewords with 170 bits are generated by an external DAC board.

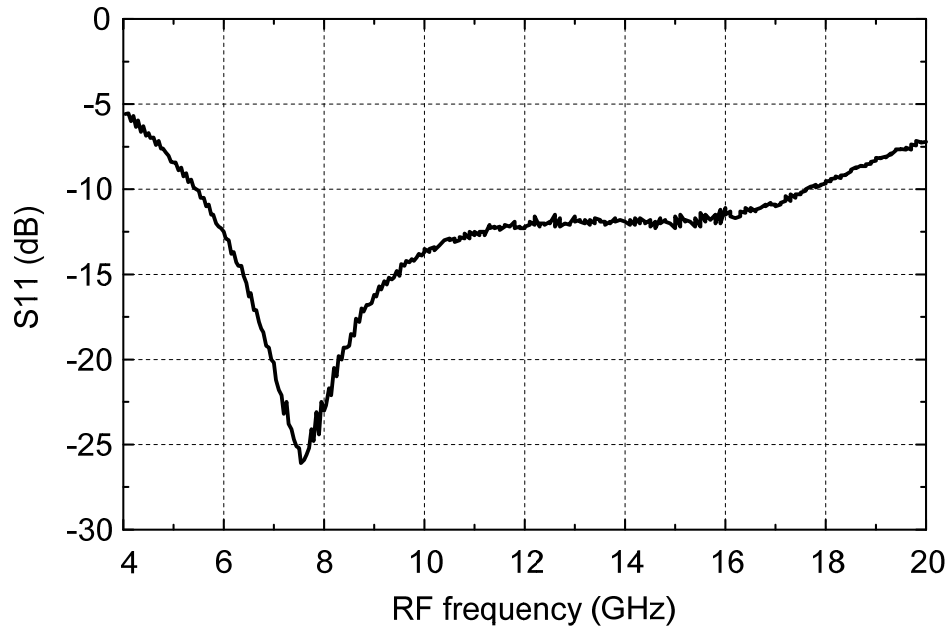


Figure 5.18: Measured input-matching performance

The RF input return loss is better than 9.8 dB across the entire band, as shown in Figure 5.18. The input-matching performance does not vary with different LC load settings of the TCA, due to the high isolation between the input and the output of the cascode stage (Figure 5.12).

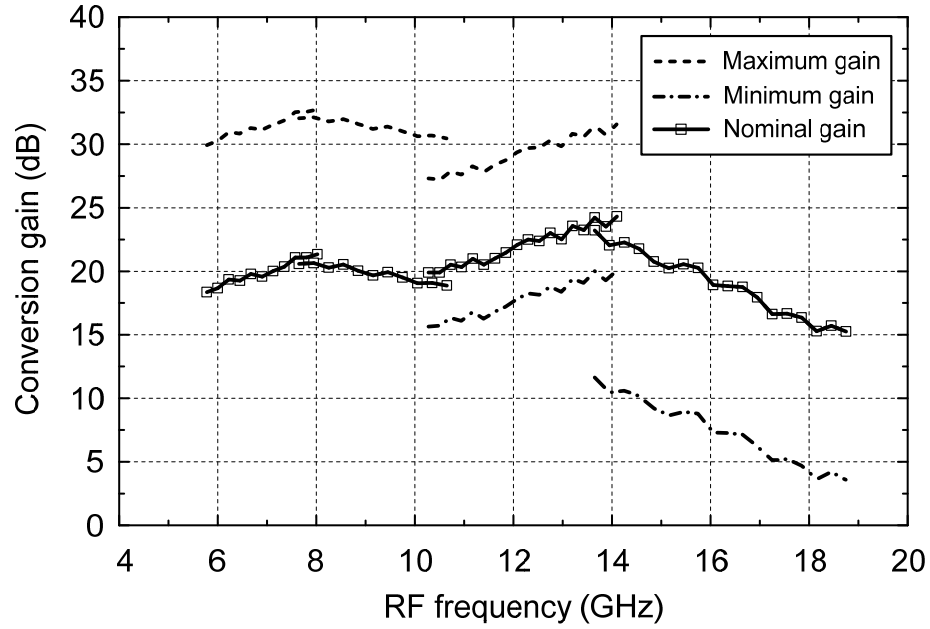


Figure 5.19: Measured conversion gain

Figure 5.19 plots the measured conversion gain of the receiver. The maximum and the minimum gains achievable are shown in dashed lines. The solid line with markers represents the nominal gain with the optimum VGA settings, which ranges from 16 to 24 dB across the entire tritave band. The discontinuities at 7.6, 10.4, and 13.5 GHz are due to the switching of either the frequency band or the IF frequency scheme.

The measured nonlinearity performance is shown in Figure 5.20. The third order intercept point (IP3) is measured by applying a two-tone signal with 10 MHz spacing. The input-referred power of third-order intercept point (IP3) and 1 dB compression does not vary with different VGA gain settings. This is because the VGA is configured by the full-scale current-commutating cells that keep the same nonlinearity performance regardless of the signal polarity.

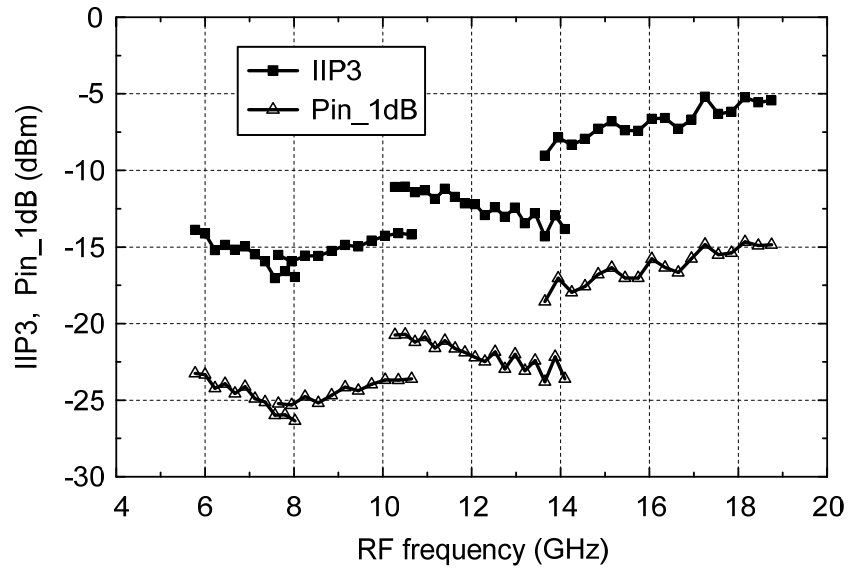


Figure 5.20: Measured nonlinearity performance: Input-referred IP3 and 1 dB compression

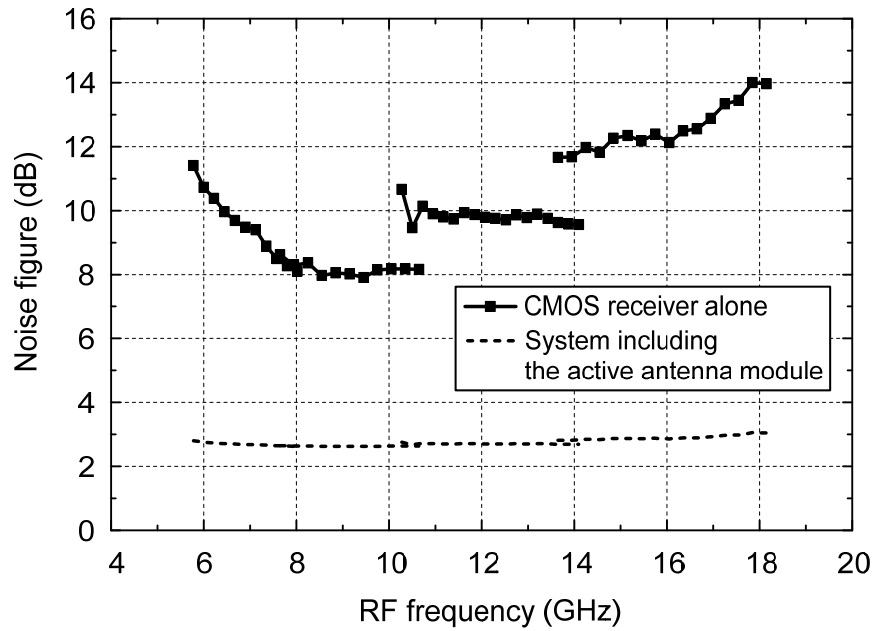


Figure 5.21: Measured noise figure of the CMOS receiver (solid line with markers) and the complete system including the active antenna module (dashed line)

The noise figure is measured by a standard Y-factor method [36]. Figure 5.21 shows the measured noise figure of the CMOS receiver, which ranges from 8 to 14 dB over the entire band. However, taking into account a preceding wideband active antenna module in the complete system (Figure. 5.3), the noise contribution of the CMOS receiver to the system will be significantly reduced. The system noise figure, including the preceding module with a 2.5 dB noise figure and a 20 dB gain, is also plotted in the dashed line.

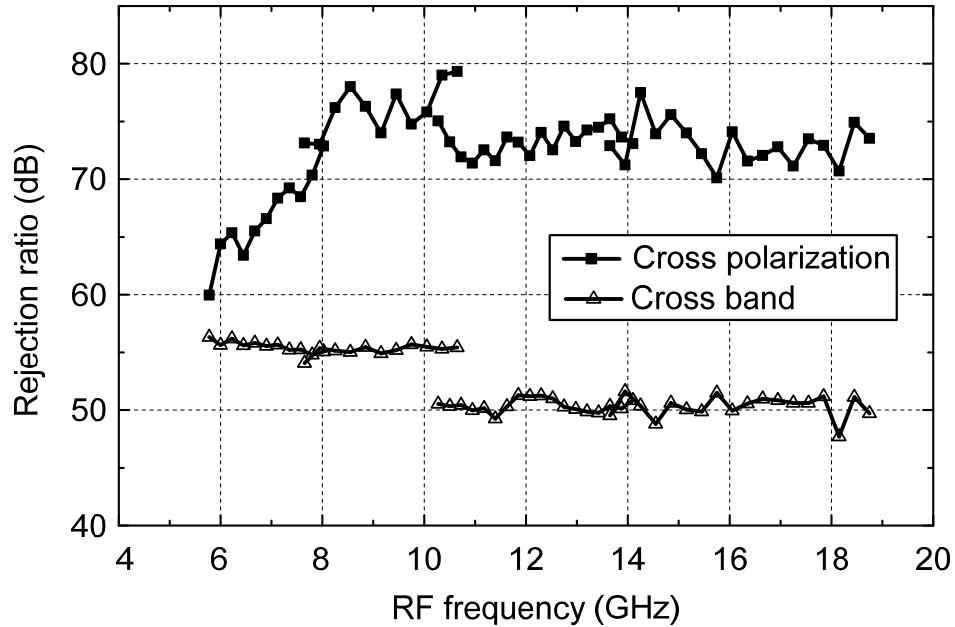


Figure 5.22: Measured isolation performance: Cross-band and cross-polarization rejection ratios

Since the receiver supports a concurrent dual-band and dual-polarization signal, it is very important to characterize the isolation performance between the two bands and between the two polarizations. For the isolation measurement, a rejection ratio is defined as a ratio of the undesired signal power, which is cross-coupled from different bands or

polarizations, to the desired signal power at the output port. For example, in order to measure the cross-band rejection ratio at the LB output port, a two-tone signal containing one LB tone and one HB tone is applied with the same input power level. Then, the rejection ratio of the HB output (the undesired cross-coupled output) is measured with reference to the LB output (the desired output) at the LB output port. As shown in Figure 5.22, the cross-band rejection ratio is more than 48 dB across the entire band. The cross-polarization rejection ratio is even better, 63 dB in the worst case.

5.5.2 Four-Element Phased Array Pattern

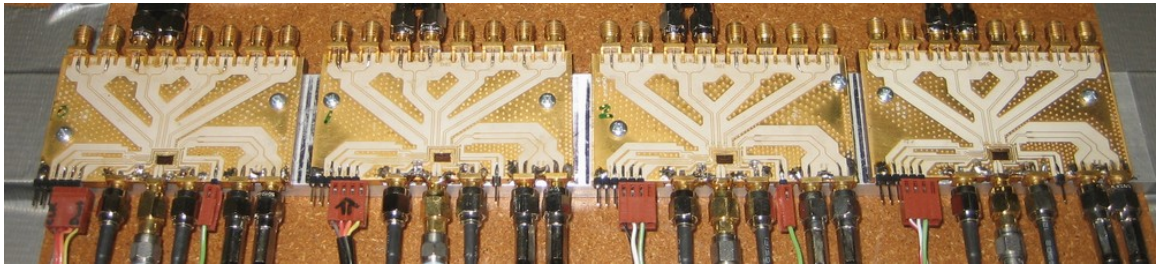


Figure 5.23: Photo of the four-element array

A four-element phased array receiver system is built by employing and incorporating four CMOS receiver chips. A photo of the module array is shown in Figure 5.23. Figure 5.24 shows the array test setup for the array pattern measurement. To characterize the array performance of the proposed system architecture alone, an electrical way of testing the array is adopted rather than using antenna modules. The RF signal generated from a signal generator is split into four RF paths by a power splitter. Each RF path feeds each array receiver element through an external variable phase shifter. By applying relative phase difference to each RF path, the external phase shifters emulate the incoming RF

wavefront at a certain incident angle. The differential baseband output from each element is converted to a single-ended signal for measurement purposes and then monitored by a four-channel digital oscilloscope. The oscilloscope performs the ideal signal combining of four channels (or four elements) internally. Alternatively, a four-way 0° power combiner is also used to combine the four baseband output signals for other measurements, such as digital demodulation or interference rejection.

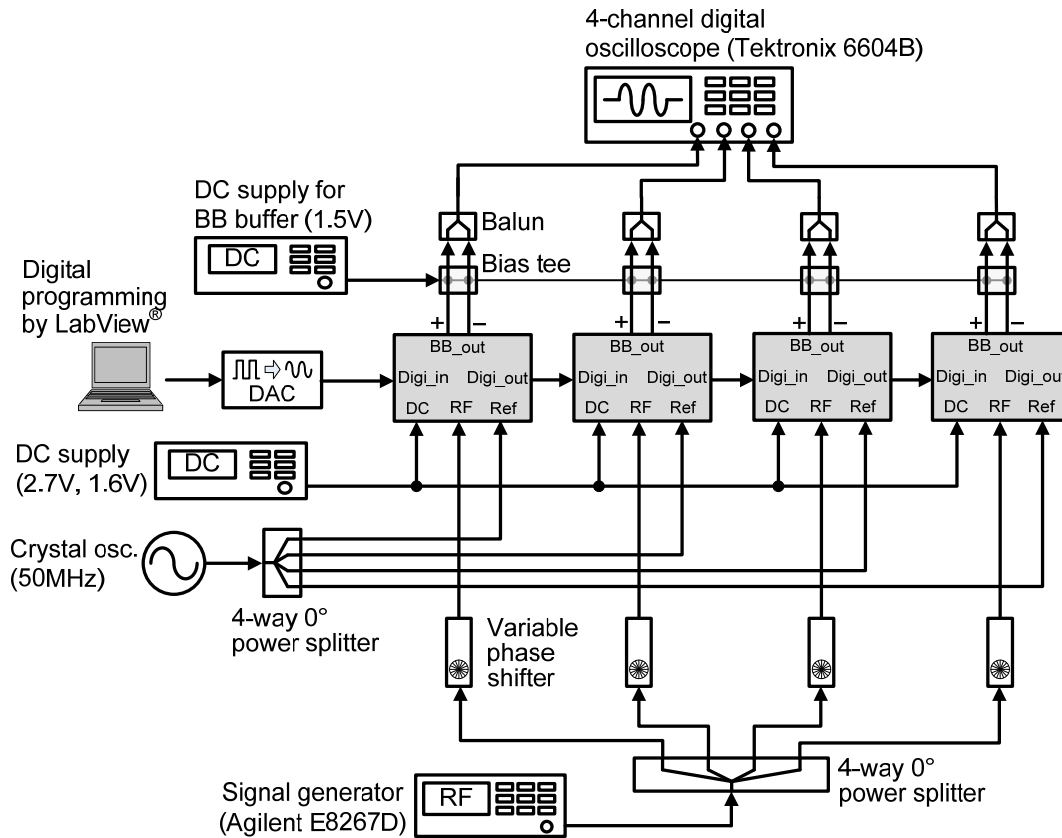


Figure 5.24: Electrical array test setup

The 50 MHz LO reference signal is generated by a crystal oscillator and then distributed to each element. A DAC board controlled by a PC generates a digital

codeword that selects the receive RF frequencies, LO frequencies, baseband gains, and the phase and amplitude interpolation.

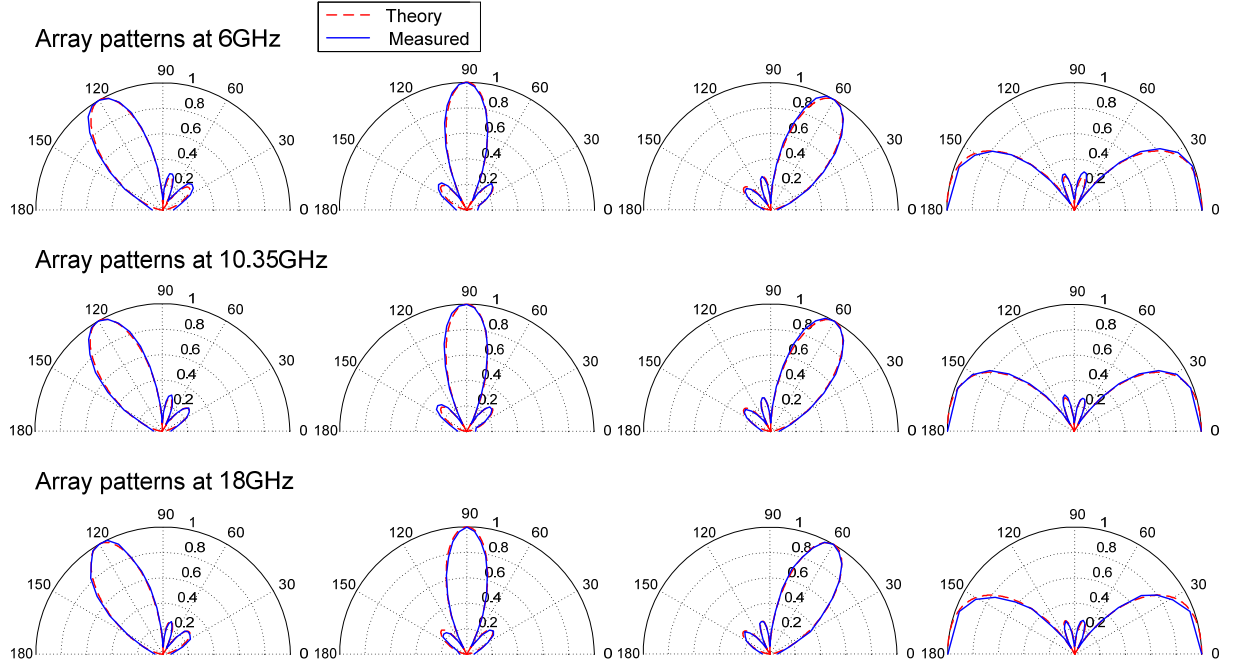


Figure 5.25: Measured array patterns of the four-element array. Theoretical patterns are superimposed.

The measured array patterns at 6, 10.35, and 18 GHz are shown in Figure 5.25. Four different beam-pointing angles are set at each different RF frequency. Theoretical patterns are superimposed on the measured ones. It can be seen that the measured beam patterns are well steered in excellent agreement with the theoretical ones. The worst-case peak-to-null ratio is 21.5 dB. This good array performance is attributed to the fine resolution of the on-chip phase shifting that enables a precise digital array calibration. The calibration offsets the process variation between different element chips and the inevitable systematic skews in phase and amplitude originating from the reference and

RF signal distribution to each array element. It also should be noted that the beam-pointing angle of the array can be steered with high resolution over the entire direction (the incident angle between -90° and 90°) due to the low RMS error of the on-chip phase shifting (see Table 5.1).

Receiver Element Performance

Conversion gain (6 – 18GHz)		16.3 ~ 24.3dB
Input-referred 1-dB compression (6 – 18GHz)		-26.3 ~ -14.8dBm
Input-referred IP3 (6 – 18GHz)		-17.0 ~ -5.2dBm
Input return loss (6 – 18GHz)		> 9.8dB
Cross-polarization rejection (6 – 18GHz)		> 63.4dB
Cross-band rejection (6 – 18GHz)		> 48.8dB
LO leakage (6 – 18GHz)		< -24.5dBm
Antenna-to-baseband noise figure [†] (6 – 18GHz)		2.6 ~ 3.1dB
RMS Phase-shifting error (6 – 18GHz)		< 0.5° (within 0.4-dB RMS amplitude variation)
RF channel spacing		225MHz (Div8 LO ₂), 300MHz (Div2 LO ₂)
Power consumption	RF and LO circuitry	658mA @2.7V, 217mA @1.6V
	Baseband buffers	34mA @1.5V each buffer
Technology		130nm CMOS
Die area		3.0×5.2 mm ²

[†]Including the active antenna module in the system.

Phased-Array Performance (four elements)

Number of beams concurrently receivable	4
Phase shifting resolution per element (6 – 18GHz)	Continuous with 0.5° RMS phase error max.
Total phased-array gain (6 – 18GHz)	28.3 ~ 36.3 dB
Beam-forming peak-to-null ratio	> 21.5dB

Table 5.1: Measured performance summary of the scalable concurrent dual-band phased array receiver

5.6 Summary

In this chapter, we introduce the basic concept of phased array, and explain that previous works on phased array systems have limited scalability and system diversity. We propose a scalable tunable dual-band quad-beam phased array receiver architecture, in which scalability is achieved with on-chip precision phase-frequency synthesizer to synchronize and calibrate the RF phase error due to reference LO routing. Difficulty for the conventional dual-band LNA in achieving required tunability has also been analyzed. As a result, the TCA is proposed to solve the issue. Four different topologies of TCA have been studied, i.e., the CG-CG, CG-CS, resistor-termination and the active termination TCAs, and we show that the active termination TCA has the largest DR. Based on the circuit studies, a phased array receiver IC is implemented in a 130 nm CMOS process for proof-of-concept, with RF measurements to verify the receiver design.

For a demonstration of the array performance, a four-element phased array system is implemented using four receiver chips. Owing to the fine resolution of the on-chip phase shifting and precise digital calibration, we achieved array patterns that agree well with the theoretical ones. To our best knowledge, this is the first concurrent multi-band multi-beam phased array receiver in a tritave bandwidth, implemented in CMOS.

Chapter 6: Concurrent Co-Channel Dual-Beam Phased Array Receiver

As mentioned in Chapter 5, a phased array receiver achieved spatial filtering property through combining the delay/phase compensated signals from different RF receiver elements. Incoming signal from a designated angle is added coherently at the baseband output. However, signals from all other directions are suppressed substantially, and information from these directions is lost at the output.

It should be noted that, though information from unwanted incoming angles are lost at the output of the phased array receiver due to spatial filtering property, information remains intact before the addition of signals from several RF elements. If we can share the RF frontend, however, reuse the amplified RF signals, feed them to several parallel phase shifter/rotator paths, and add the corresponding parallel phase-compensated outputs concurrently, we can achieve concurrent multi-beam in a single phased array receiver system. Compared to conventional approaches, a concurrent multi-beam phased array reduces the number of frontend antennas and RF circuits. As a result, complexity and cost of multi-beam phased array systems can be minimized.

In Section 6.1, we will introduce the system architecture of the proposed 10.4–18 GHz co-channel dual-beam CMOS receiver. Section 6.2 will discuss the architecture of the receiver element, and the circuit implementations of the RF/IF/baseband signal path. Section 6.3 presents the experimental results on the receiver element, and a four-element

co-channel phased array receiver system. Finally, a summary is provided in Section 6.4.

6.1. Dual-Beam Phased Array System Architecture

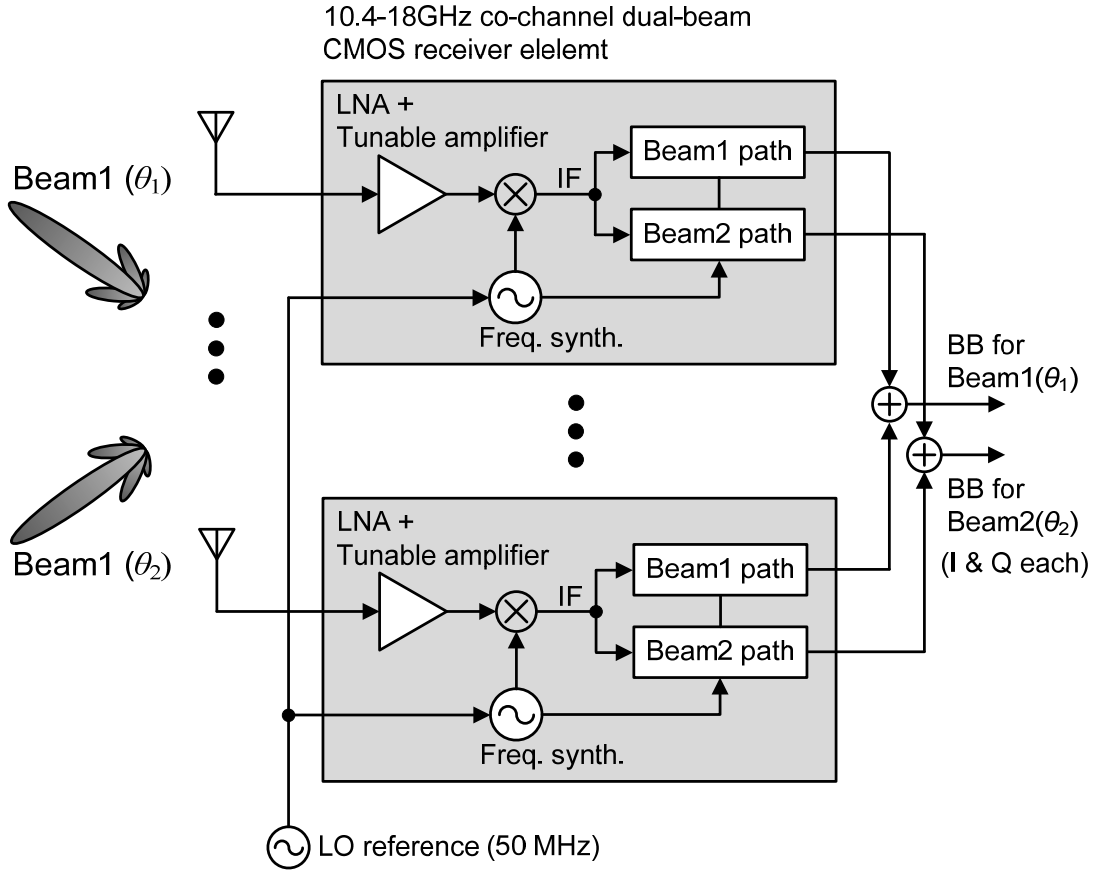


Figure 6.1: Architecture of the 10.4 – 18 GHz co-channel dual-beam phased array receiver system

A co-channel dual-beam phased array system concurrently forms two independent beams with different spatial signatures at a single RF frequency (or channel) between 10.4–18 GHz. As shown in Figure 6.1, two incoming co-channel signals that have different spatial signatures, e.g., different DOAs, are received by the antenna array. Each

antenna feeds a single CMOS phased array element that performs quadrature down-conversion for each beam signal with independent phase shifting. The receive frequency of the CMOS receiver element is tunable between 10.4 and 18 GHz. The co-channel signals are amplified, filtered, and down-converted to the IF by a LNA, a tunable amplifier, and an RF mixer, respectively. Subsequently, the IF signal is split into two paths. Each path corresponding to each beam takes separate phase shifting and down-conversion to different baseband outputs. The quadrature baseband outputs from multiple CMOS chips are combined off-chip to complete the beamforming. The LO signals are generated from an on-chip frequency synthesizer with a 50 MHz reference signal supplied by an off-chip precision crystal oscillator.

It should be noted that the quad-beam system introduced in Chapter 4 can also be reconfigured to receive two co-channel beams through the two input ports for different polarizations (HP and VP in Figure 5.3). However, it requires splitting the signal received at each antenna into two replicas before feeding to the two inputs (HP, VP) of the CMOS receiver element. Therefore, such an approach will lead to at least a 3 dB loss of gain and noise figure ahead of the receiver element, and result in degradation of the receiver sensitivity. Furthermore, since two physically different TCAs are used to amplify the co-channel signals at the same carrier frequency, the power and area consumption will be unnecessarily doubled.

The co-channel dual-beam system can also be reconfigured as a smart-antenna system to take advantage of the spatial division multiple access (SDMA), spatial multiplexing (SM) techniques, as well as the spatial diversity. ADCs and back-end DSP units will be required

to adaptively control the phase and amplitude of each element during the dual beamforming operations [44] [45]. Since the beamforming in the proposed system is performed in the analog domain, the processing speed required in the DSP units will be substantially relieved [45] [46]. This allows for high-throughput real-time beamforming without experiencing severe I/O data congestion in the back-end units.

Finally, it should be emphasized that the proposed system is easily scalable to implement a large-size array. More CMOS elements can be added in the system with a relatively small extra cost. Also, since each element output and the LO reference signal is combined/distributed at the low frequency, the proposed system avoids the complicated network that would be required to combine/distribute high-frequency signals in a conventional large-size array.

6.2. A 10.4–18 GHz Concurrent Quad-Beam Phased Array Receiver

6.2.1. Receiver Element Block Diagrams

The architecture of the 10.4–18 GHz co-channel dual-beam receiver element is shown in Figure 6.2. The incoming dual-beam signal (Beam 1 and Beam 2) between 10.4 and 18 GHz is received and amplified by a front-end LNA. A subsequent tunable amplifier attenuates out-of-band frequency components before down-conversion. It should be noted that the two distinct beams contained in the incoming signal share the same RF channel with different spatial signatures. This means that they are indistinguishable from each other

before the beamforming operation. Therefore, a single set of the LNA and the tunable amplifier amplifies the two beams concurrently and presents only a single output. This is contrary to the TCA in Section 5.3 which provides dual outputs for two different bands.

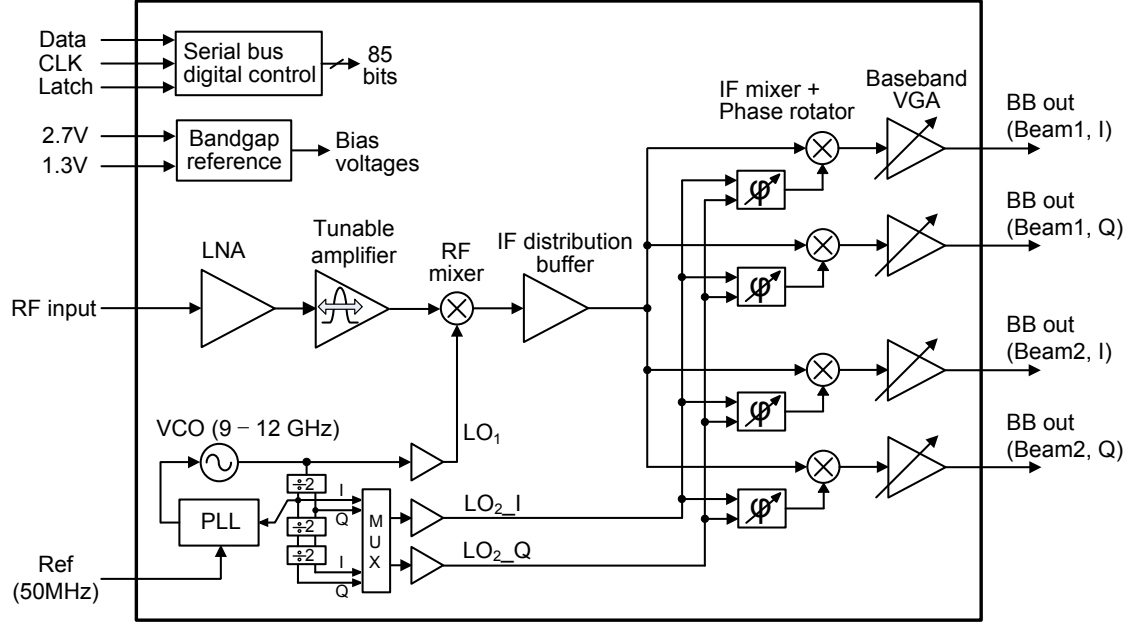


Figure 6.2: Architecture of the tunable co-channel dual-beam phased array receiver element in CMOS (10.4–18 GHz).

The output signal of the tunable amplifier is then down-converted by an RF mixer and finally split into two separate beam paths by an IF distribution buffer for independent beamforming. Each beam path has two sets of an IF mixer and a baseband VGA for quadrature down-conversion. Consequently, the IF distribution buffer drives four IF mixers in total.

The LO signal driving the RF and IF mixers is generated by an on-chip frequency synthesizer. The VCO provides the RF mixer with the first LO (LO₁) frequency between 9

and 12 GHz. The second LO (LO2) frequency for the IF mixers is selected between one eighth (1.125–1.5 GHz) and one half (4.5–6 GHz) of the LO1 frequency by a multiplexer. Accordingly, any RF frequency between 10.125 and 18 GHz is down-converted to baseband without any blind spot by selecting one of the two LO2 frequency bands.

For receiver beamforming capability, the phase of the LO2 signal is shifted by a 10 bit digital phase rotator. Each IF mixer is driven by a separate phase rotator for the independent beamforming operation.

6.2.2. LNA

In addition to the amplification of the weak incoming RF signal, the LNA provides a wideband input matching to $50\ \Omega$ and conversion of the single-ended input to a differential output. In Figure 6.3, the input signal feeds the cascode (M1 and M2) and the common gate (M3) stages in parallel. By feeding the output of the cascode back to the gate of M3, the gm of M3 is boosted effectively by a factor of $(1+A_c)$, where A_c is the cascode gain [47]. The gm boosting enables the common gate to provide a low input impedance for the input matching with less bias current, which generates less channel noise accordingly. The residual input reactance is resonated out with a LC network to achieve the wideband matching to $50\ \Omega$. In addition, the LNA operates as an active balun by taking out-of-phase output signals from the cascode and the common gate, respectively. An inductor (0.71 nH) in parallel with a resistor (150 Ω) is used to compensate for the phase delay introduced by parasitic capacitances in the cascode path.

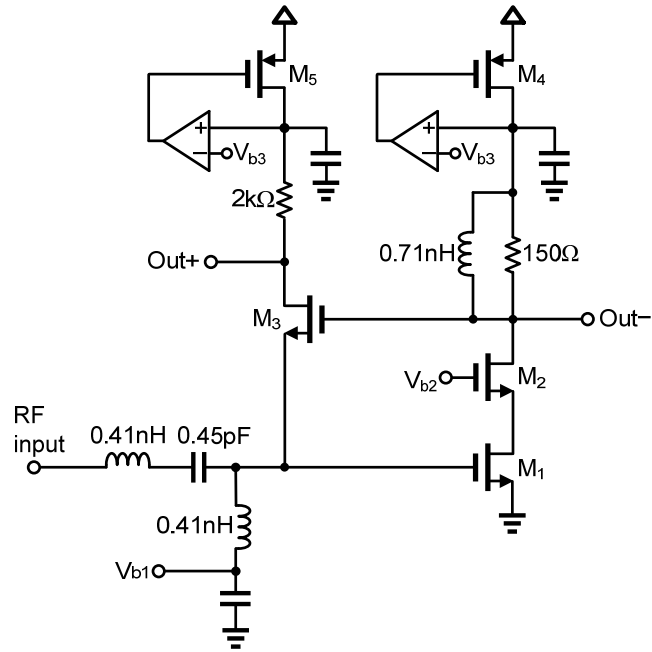


Figure 6.3: Schematic of the LNA

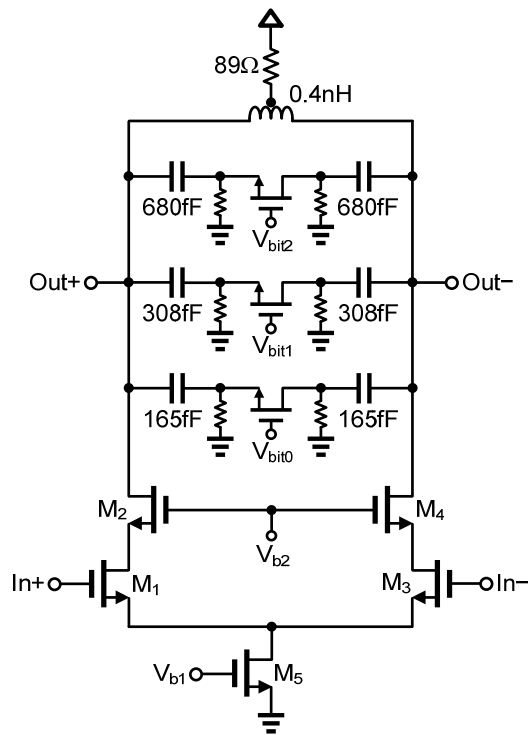


Figure 6.4: Schematic of the tunable amplifier

The tunable amplifier shown in Figure 6.4 is configured as a differential cascode amplifier (M1–M4). More gain is added to the LNA output signal in order to reduce noise contribution of the subsequent mixer blocks. The center frequency of the amplifier is controlled by the 3-bit switched capacitors in the output LC tank. Out-of-band signals are attenuated according to the tuned gain performance.

6.2.3. IF Signal Distribution Networks

In the dual-beam receiver architecture, IF and LO signals need to be distributed to four different IF mixers through relatively long distribution paths and large node parasitic capacitance. The distribution buffers use a wideband Cherry-Hooper topology to eliminate the large shunt-peaking inductors that were employed in the 6–18 GHz quad-beam receiver element given in Chapter 5. This reduces power consumption and chip area substantially while achieving the required wideband operation.

6.2.4. RF/IF Mixers and Baseband Buffers

The current-commutating double-balanced mixers are implemented for the RF and IF mixers. The output of the RF mixer is configured as a tunable LC tank similar to that of the tunable amplifier in Figure 6.4. The IF mixers use RC loads in the baseband outputs.

6.2.5. Receiver Element Implementation

The tunable co-channel dual-beam phased array receiver elements are implemented in a 130 nm CMOS process. The die micrographs are shown in Figure 6.5, and the chip area is

$2.1 \times 1.6 \text{ mm}^2$. This chip occupies much less area than the quad-beam receiver described in Chapter 5, with a 78 % reduction. This is not only due to the smaller number of beams to be supported, but also due to the inductorless design of the IF and LO buffers and phase rotators.

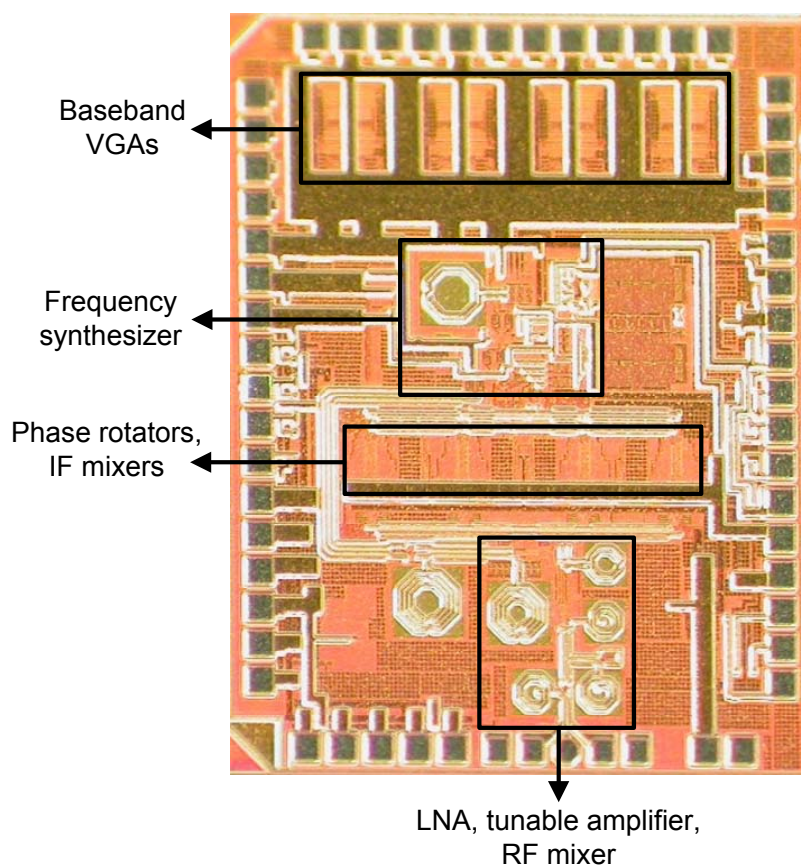


Figure 6.5: Chip micrograph of the 10.4 –18 GHz dual-beam receiver element

6.3. Experimental Results

6.3.1. Receiver Element Measurement Results

The 10.4–18 GHz dual-beam receiver chip is mounted and wire-bonded on a testing module for the measurements, similar to the quad-beam receiver element setup. The measured conversion gain and input-matching performance are shown in Figure 6.6. The conversion gain ranges from 22 to 27 dB over the entire band of 10.4–18 GHz. The discontinuity at 13.5 GHz is due to the LO₂ frequency switching (Section 6.2.1). The input reflection coefficient is lower than –10.4 dB over the entire band.

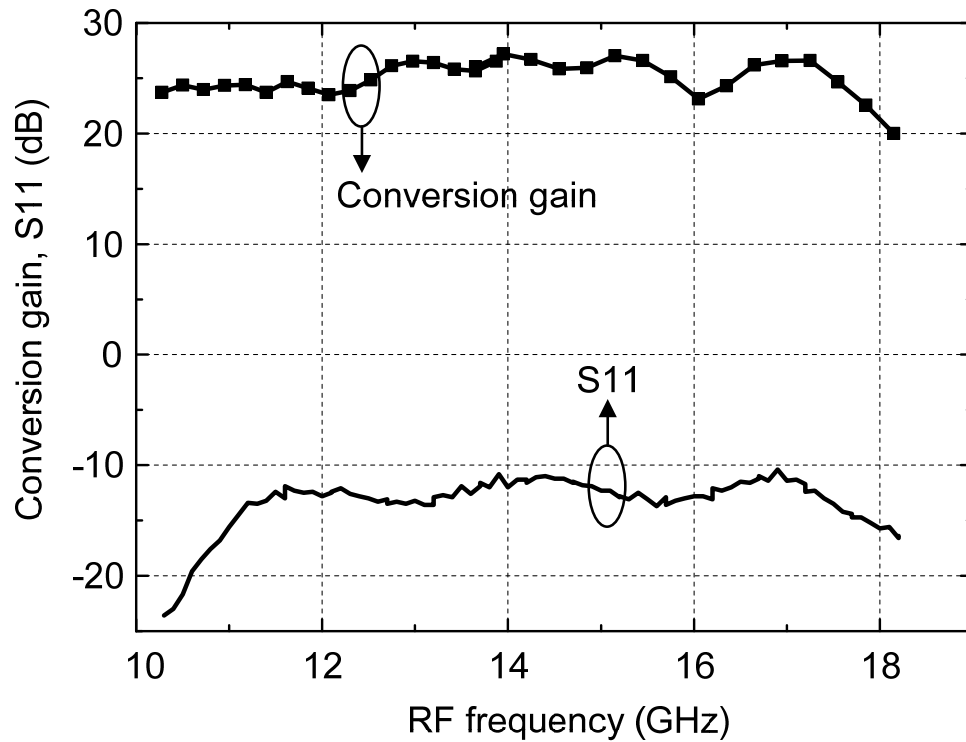


Figure 6.6: Measured conversion gain and input-matching performance of the 10.4–18 GHz dual-beam receiver element

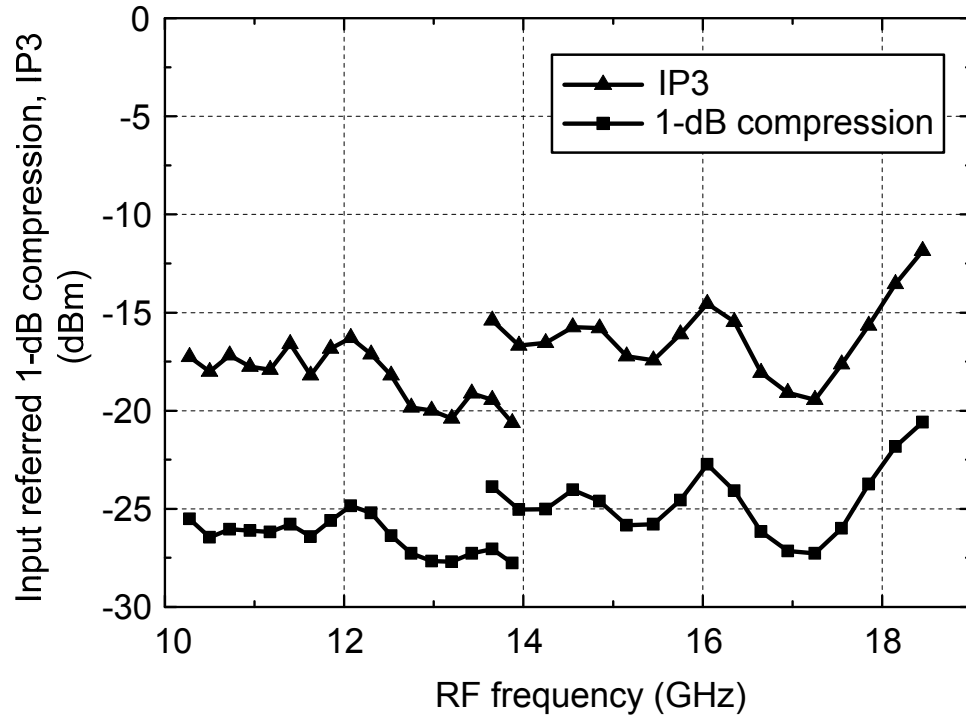


Figure 6.7: Measured input-referred 1 dB gain compression and the IP3 of the 10.4–18 GHz dual-beam receiver element.

The measured 1 dB gain compression and IP3 performance are shown in Figure 6.7. The input-referred 1 dB compression power is less than that of the 6–18 GHz quad-beam receiver element, because the conversion gain is improved in this dual-beam receiver element. Figure 6.8 shows the measured noise figure, which ranges from 4.4 to 9.5 dB. Due to its single-band operation, the noise figure is improved by 4 dB nominally, compared to the 6–18 GHz quad-beam receiver element.

The receiver element chip draws 225 mA and 74 mA from DC supplies of 2.7 V and 1.3 V, respectively. The power consumption is significantly improved, compared to the quad-beam receiver element. Each baseband buffer draws 34 mA from a 1.5 V supply.

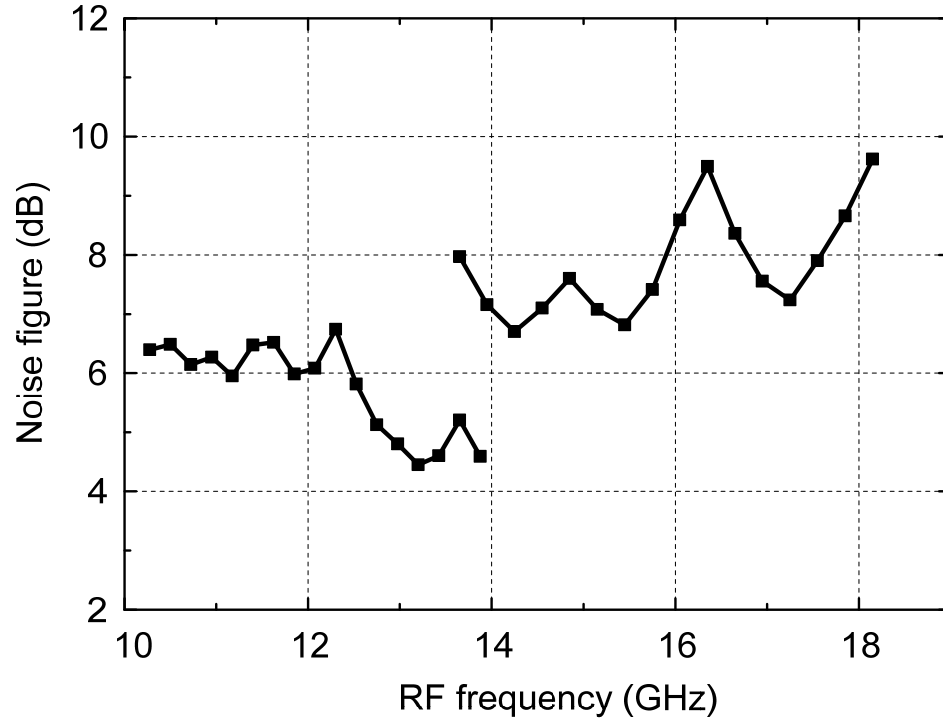


Figure 6.8: Measured noise figure of the 10.4–18 GHz dual-beam receiver element

6.3.2. Four-Element Phased Array Measurement Results

To verify the co-channel dual-beamforming capability in the 10.4–18 GHz, two RF signals are generated by two signal generators, respectively, each fed to the array with different DOAs (Figure 6.9). One signal (denoted as beam 1) is split by a four-way power splitter and fed to each element through identical fixed-phase paths, emulating a fixed normal incidence ($\theta = 0^\circ$). The other signal (denoted as beam 2) is also split into four paths, and the phase of each path is shifted by an external variable phase shifter, emulating an arbitrary incident angle. The two signals are combined by a two-way 0° power combiner to feed each element. The rest of the test setup is similar to that of Figure 5.23.

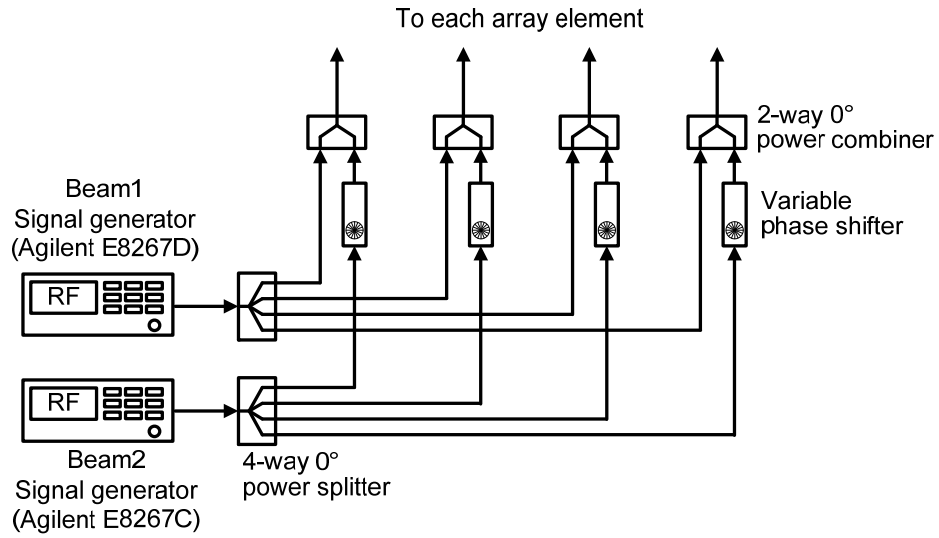
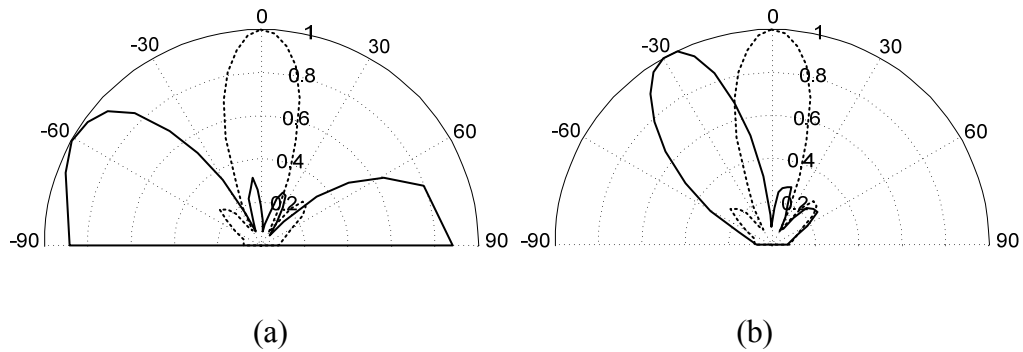


Figure 6.9: Concurrent co-channel dual-beam feed with different DOAs

The measured electrical array patterns are shown in Figure 6.10 for the concurrent dual beamforming at 17.85 GHz. Each beam is measured at the corresponding separate output ports of the array. The beam-pointing angle for beam 2 is steered at -60° , -30° , 30° , and 60° , respectively, while the beam-pointing angle for beam 1 is fixed to 0° . The dual beamforming is successfully achieved with the accurate beam-steering operation. This demonstrates the system capability for the SDMA and SM techniques when the elements are controlled adaptively. Similar results are obtained for other RF frequencies including 10.5 and 13.95 GHz.



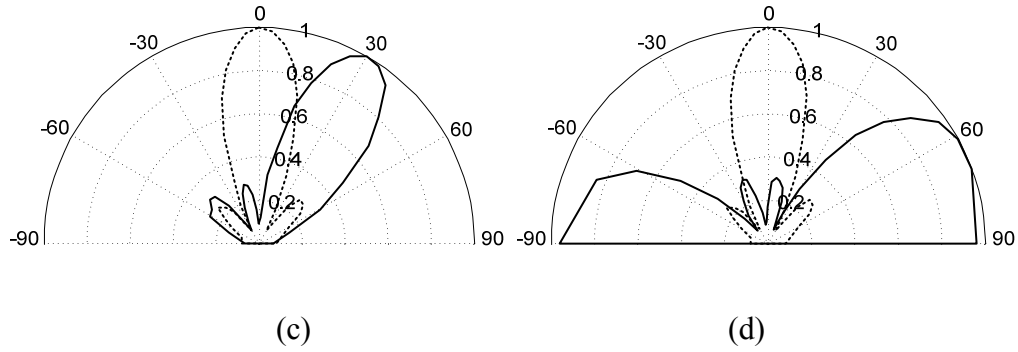


Figure 6.10: Measured concurrent dual-beam array patterns at 17.85 GHz of the 10.4–18 GHz co-channel dual-beam phased array. The beam-pointing angle for beam 1 (dashed line) is fixed at 0°. The beam-pointing angle for beam 2 (solid line) is steered at (a) –60°, (b) –30°, (c) 30°, (d) 60°. The antenna spacing is assumed as a half wavelength of the incoming signal.

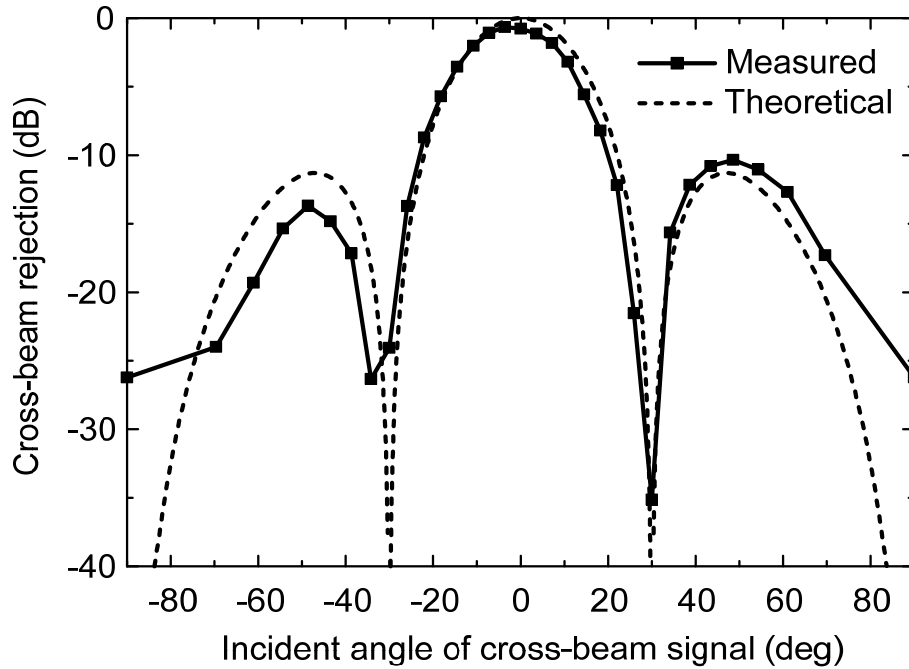


Figure 6.11: Measured cross-beam rejection performance ($f_{RF} = 17.85$ GHz). The incident angle of the desired signal is fixed at 0°.

The cross-beam rejection ratio is measured in a similar way to the dual-beam array pattern measurement. Two CW signals at 17.85 and 17.86 GHz are generated with an identical power level of -36 dBm and concurrently applied to the array with different DOAs. The resulting two output signals from the array are measured together at the same output port of beam 1. Thus, the rejection ratio of beam 2 (undesired cross-coupled signal) with respect to beam 1 (desired signal) is measured as a function of the incident angle of beam 2. To distinguish the two beams at the same output port, a small frequency spacing (10 MHz) is applied between the two CW signals. The measured rejection performance is shown in Figure 6.11. As expected, the measured rejection curve follows well the theoretical one calculated assuming the ideal combining with equal amplitudes. This verifies the established spatial filtering performance of the phased array system. The rejection ratio at the null positions ($\pm 30^\circ$ and 90°) is better than 24 dB. The beamwidth for 10 dB rejection is 44° . It should be noted that if more elements (N) are combined, the number of null positions will increase to $(N-1)$ and the beamwidth will be narrower.

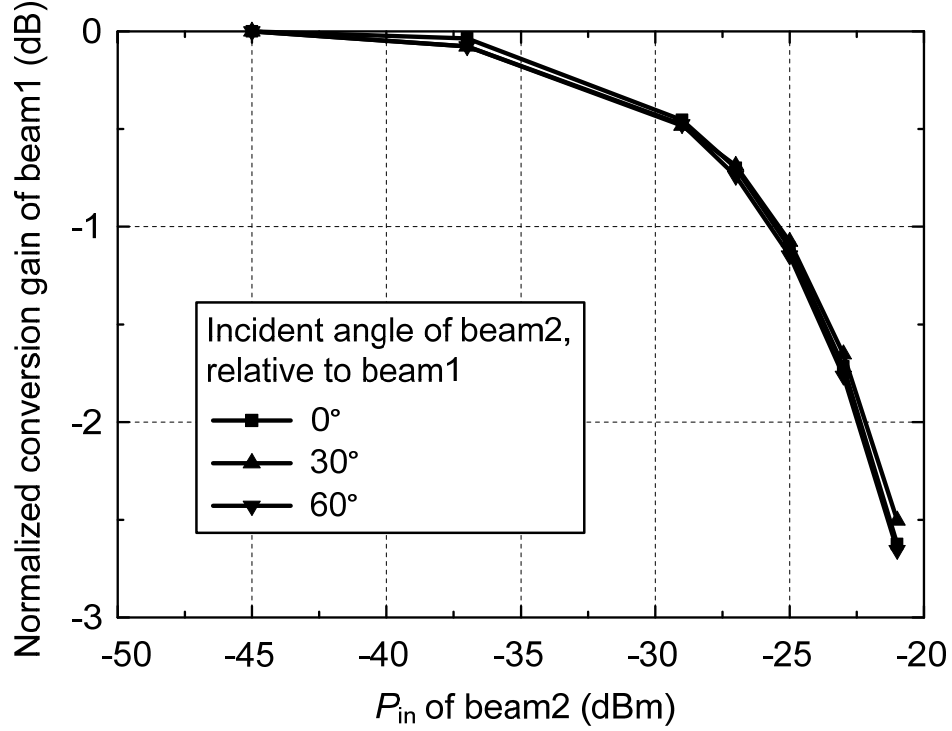


Figure 6.12: Desensitization of the array system ($f_{RF} = 17.85$ GHz)

If two beams arrive at the array with substantially different power levels, the beam with strong power will desensitize the receiver and reduce the gain of the other beam. In principle, assuming a memoryless and time-invariant system, the cross 1 dB gain compression due to a strong interference is 3 dB less than the regular 1 dB gain compression due to a single-tone signal [48]. To characterize the desensitization of the array system, two beams (beam 1 and beam 2) at 17.85 and 17.86 GHz are applied with different power levels. Beam 1 is set to -45 dBm and beam 2 is swept from -45 to -21 dBm. The array output power of beam 1 is then measured (Figure 6.12). The conversion gain of beam 1 is reduced by 1 dB at -25.3 dBm of the input power of beam 2. It is compared with -23.7 dBm of the input-referred 1 dB gain compression measured at the

CMOS receiver element at the same frequency (Figure 5.7). It can also be seen in Figure 5.12 that the 1 dB compression by the strong interferer does not depend on the incident angle of the interferer. This is because the desensitization occurs in the receiver block of each array element before beamforming.

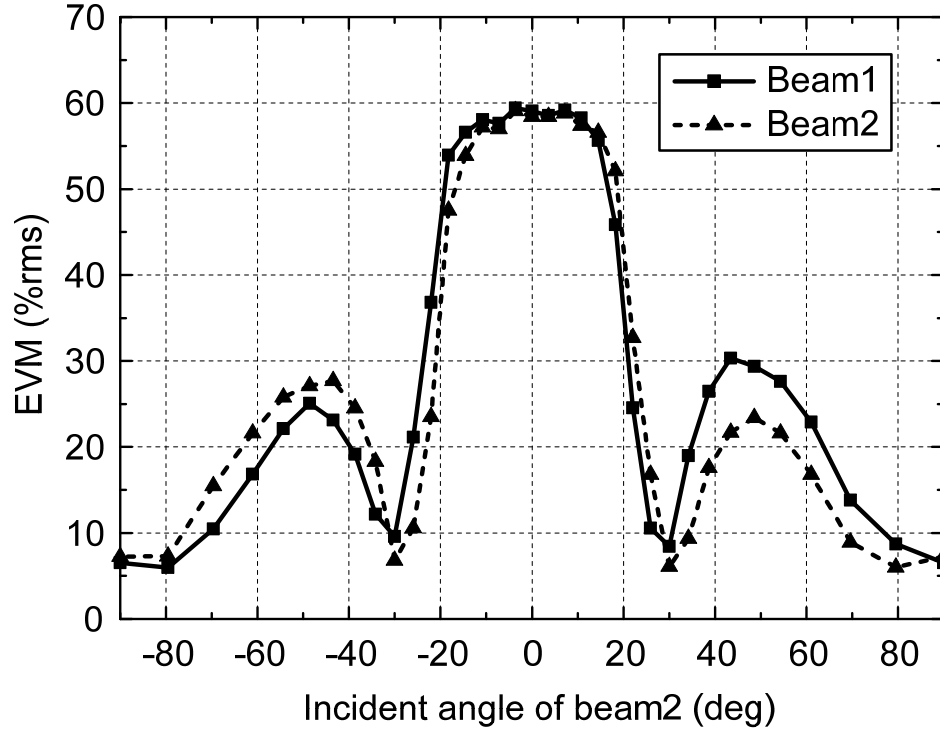


Figure 6.13: Measured EVM of the concurrent dual-beam signals, each independently modulated with 4.5 Msps QPSK at 17.85 GHz. The incident angle of beam 1 is fixed at 0°.

Finally, the digital demodulation performance of the concurrent two beams is measured to further demonstrate the SDMA and SM capability of the proposed multi-beam system. Two RF signals, each independently modulated by 4.5 Msps QPSK at 17.85 GHz with an identical power level, are fed to the array using the setup shown in

Figure 6.9. The incident angle of beam 2 is swept from -90° to 90° while beam 1 arrives with a fixed incident angle of 0° . The EVM is measured at the two array output ports for beam 1 and beam 2, respectively. As shown in Figure 5.13, the EVM of both beams decreases rapidly as the two beams are separated FROM each other spatially. At the null positions ($\pm 30^\circ$ and 90°), the two beams do not interfere with each other and recover a small EVM. Two final notes about the measurement result are given as follows: First, the measured EVM is much higher than the actual EVM of the array system due to substantial loss of 39 dB in the test-purpose RF distribution network (Figure 6.9). Second, the EVM will be reduced more sharply in the spatial domain with a larger array size, which can be achieved by the proposed systems with low cost and complexity. The performance summary of the concurrent co-channel dual-beam receiver element and phased array receiver system is shown in Table 6.1.

6.4. Summary

A concurrent co-channel dual-beam receiver that receives two beams at the same frequency between 10.4–18 GHz is proposed and implemented in a 130 nm CMOS process. Due to the scalable system architecture and the integration of array-receiver components in a single chip, a large number of array elements can be added to build a very large-scale array with low cost and complexity. For demonstration purposes, a four-element phased array system has been implemented to verify the benefits of dual-beam phased array approach. To the best of our knowledge, this is the first CMOS-based array system that supports co-channel dual-beam over the wide frequency range.

Receiver Element Performance

Conversion gain (10.4–18 GHz)		22~27 dB
Input-referred 1 dB compression (10.4–18 GHz)		-28~-23 dBm
Input-referred IP3 (10.4–18 GHz)		-21~-15 dBm
Input return loss (10.4–18 GHz)		> 11 dB
LO leakage (10.4–18 GHz)		< -23 dBm
Receiver element noise figure (10.4–18 GHz)		4.4~9.5 dB
RMS phase-shifting error (10.4–18 GHz)		< 0.6° (within 0.4 dB RMS amplitude variation)
RF channel spacing		225 MHz(Div8 LO ₂), 300 MHz (Div2 LO ₂)
Power consumption	RF and LO circuitry	225 mA@2.7 V, 74 mA@1.3 V
	Baseband buffers	34 mA @1.5 V each buffer
Technology		130 nm CMOS
Die area		2.1×1.6 mm ²

Phased Array Performance (four elements)

Number of beams concurrently receivable	2
Phase shifting resolution per element (10.4–18 GHz)	Continuous with 0.6° RMS phase error max.
Total phased array gain (10.4–18 GHz)	34~39 dB
Beam-forming peak-to-null ratio (17.85 GHz)	> 24 dB

Table 6.1: Measured performance summary of the concurrent co-channel dual-beam phased array receiver

Chapter 7: Conclusion

This thesis presents a study of circuits and systems of wireless concurrent communication. The contributions of our study include the development of original concepts and new theoretic findings together with practical implications. As a result, integrated wireless systems with more link diversity and data rate have been devised.

7.1 Summary

We have presented a unique view on wireless radio frontend systems that use concurrency for analog signal processing. Concurrency is a special kind of circuit parallelism that uses a single circuit with necessary bandwidth to process multiple signals at the same time. Concurrent radios offer a higher data-rate and improved system diversity. Our comprehensive treatment is comprised of proposals for potential transceiver architectures, invention of circuit blocks, and provisions of innovative analysis methods.

The analysis of concurrent circuits are often complex. To simplify noise analysis, we proposed a R^{N^2} -vector space for modeling an arbitrary noisy network, and proved that any internal physical sources inside the noisy network contributes a small vector in the defined R^{N^2} -vector space. The aggregate statistical behavior of this noisy network can be viewed as the vector sum of these vectors. A general two-port noisy network has been demonstrated as an example. Its application to modeling FETs leads to several modified FET noise models, in which three uncorrelated noise sources are sufficient to describe the

statistical behavior of an intrinsic FET. The use of these new FET models can simplify the analysis, simulation and optimization of low-noise systems without sacrificing accuracy.

Broadband low-noise amplifier is a critical block in concurrent receiver systems. We first reviewed the Bode-Fano criteria and discussed the matching, noise figure, and power trade-offs in designing a conventional broadband CMOS LNA. To deal with this trade-offs, we propose a novel low noise weighted distributed amplifier (WDA) topology, which uses the internal finite-impulse-response filtering inside a conventional distributed amplifier to partially suppress internal thermal noise. A distinct advantage of this topology is its tolerance to input parasitic capacitance which can be used to provide good electrostatic discharge (ESD) protection without sacrificing its noise performance and power consumption. The proposed modified FET noise model is used to simplify WDA's analysis and optimization. A 3.1–10.6 GHz WDA is implemented on a 130 nm CMOS process. The use of alternating coupling LC-ladders further shrinks chip size to a compact $870 \times 500 \mu m^2$ area. Experimental results show 2.3–4.5 dB NF at 23 mW power consumption.

Using concurrency in wireless links can boost communication data rate. As a proof-of-concept, we proposed dynamic scalable concurrent communication by dividing the 7.5 GHz bandwidth of the unlicensed 3.1–10.6 GHz spectrum into seven concurrent channels. A CMOS octa-core RF receiver was implemented and measured to demonstrate the concept. Based on the receiver measurement results, a wireless link can be built to achieve a 16 Gbps channel limit at five meter TX-RX distance at 400 mW power consumption.

Tunable concurrency can improve the receiver diversity. A prototype 6–18 GHz concurrent tunable dual-band phased array receiver element IC is proposed and built on a 130 nm CMOS process. Design challenges and proposed solutions to achieve dual-band RF signal reception have been studied. Experimental results demonstrate successful dual-band RF reception within a high band (6–10.4 GHz) and low band (10.4–18 GHz) with 300 MHz baseband bandwidth. A final four-element phase array receiver built from the prototyped ICs shows an array pattern with worst-case 21 dB peak-to-null ratio across all frequencies.

A phased array receiver presents spatial filtering property at the system output. However, it is noted that information from different incoming angles are intact before the combining of phase-compensated receiver array outputs. We have used this property to design a CMOS 10.4–18 GHz concurrent dual-beam phased array receiver. Antennas, RF frontend, and LO circuits are shared between the two beam paths to reduce overall system complexity. A prototype receiver IC is implemented on a 130 nm CMOS process. A final four-element phased array receiver shows successful concurrent dual-beam reception at the same RF frequency.

Bibliography

- [1] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, v.32, July 1928, pp. 97-109.
- [2] H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, v. 32, July 1928, pp. 110-13.
- [3] H. A. Haus and R. B. Adler, *Circuit Theory of Linear Noisy Networks*, The Technology Press of The Massachusetts Institute of Technology and John Wiley & Sons, Inc., New York, 1959
- [4] J. N. Franklin, *Matrix Theory*, Dover Publications Inc., New York, 2000
- [5] L. O. Chua, C. A. Desoer, and E. S. Kuh, *Linear and Nonlinear Circuits*, McGraw-Hill Book Company, Singapore, 1987
- [6] A. Leon-Garcia, *Probability of Random Processes for Electrical Engineering, Second Edition*, Addison Wesley Longman, New York, 1994
- [7] A. V. D. Ziel, "Thermal Noise in Field-Effect Transistors," *Proc. IRE*, v. 50, Aug 1962, pp. 1808-12.
- [8] A. V. D. Ziel, "Gate Noise in Field-Effect Transistors at Moderate High Frequencies," *Proc. IRE*, v. 51, Mar 1963, pp. 461-467.
- [9] H. Rothe, "Theory of Noisy Fourpoles," *Proc. IRE*, pp. 811-818, June 1956.
- [10] M. W. Pospieszalski, "Modeling of Noise Parameters of MESFET's and MODFET's and Their Frequency and Temperature Dependence," *IEEE Trans. on MTT.*, Vol. 37, Sept. 1989.

- [11] M. W. Pospieszalski, "On Certain Noise Properties of Field-Effect and Bipolar Transistors", *Proc. of MIKON Conference*, May 2006.
- [12] BSIM4 manual: <http://www-device.eecs.berkeley.edu/~bsim3/bsim4.html>.
- [13] J. S. Goo, *High Frequency Noise in CMOS Low Noise Amplifier*, Ph.D. thesis, Stanford University, 2001.
- [14] M. Shoji, "Analysis of High-Frequency Thermal Noise of Enhancement Mode MOS Field-Effect Transistors", *IEEE Trans. on Electron Devices*, Vol.13, Jun 1966, pp.520-24.
- [15] R. M. Fano, *Theoretic Limitations on the Broadband Matching of Arbitrary Impedances*, Massachusetts Institute of Technology, Jan. 2nd, 1948.
- [16] P. Heydari, "Design and analysis of a performance-optimized CMOS UWB distributed LNA," *IEEE JSSC*, pp. 1892-1904, Sept. 2007.
- [17] B. Analui, and A. Hajimiri, "Bandwidth enhancement for transimpedance amplifier," *IEEE JSSC*, pp. 1263-70, Aug. 2004.
- [18] M.T. Reihha, and J. Long, "A 1.2V reactive-feedback 3.1-10.6 GHz low-noise amplifier in 0.13 μ m CMOS", *IEEE JSSC*, pp. 1023-33, May 2007.
- [19] C-F. Liao, and S-I Liu, "A broadband noise-canceling CMOS LNA for 3.1-10.6 GHz UWB receivers," *IEEE JSSC*, pp. 329-339, Feb. 2007.
- [20] J. Borremans, et al, "An ESD-protected DC-to-6GHz 9.7mW LNA in 90nm digital CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 422-423, Feb. 2007.
- [21] R. Ramzan, et al., "A 1.4V 2.5mW inductorless wideband LNA in 0.13 μ m CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 424-425, Feb. 2007.

- [22] X. Guan, H. Hashemi, and A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2311–2320, Dec. 2004.
- [23] X. Guan and A. Hajimiri, "A 24GHz CMOS front-end," in *Proc. of IEEE European Solid-State Circuits Conference (ESSCIRC)*, Sep. 2002, pp. 155-158.
- [24] A. Natarajan, A. Komijani, and A. Hajimiri, "A fully integrated 24-GHz phased-array transmitter in CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2502–2514, Dec. 2005.
- [25] A. Natarajan, A. Komijani, X. Guan, A. Babakhani, and A. Hajimiri, "A 77-GHz phased-array transceiver with on-chip antennas in silicon: transmitter and local LO-path phase shifting," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2807–2819, Dec. 2006.
- [26] K. Hansen, "Wireless RF design challenges," in *Radio Freq. Integr. Circuits Symp.*, Jun. 2003, pp. 3–7.
- [27] D. Porcino and W. Hirt, "Ultra-wideband radio technology: potential and challenges ahead," *IEEE Communications Magazine*, vol. 41, issue 7, pp. 66-74, Jul. 2003.
- [28] K. Koh and G. M. Rebeiz, "An eight-element 6 to 18 GHz SiGe BiCMOS RFIC phased-array receiver," *Microwave Journal*, pp. 270-274, May 2007.
- [29] T.-S. Chu, J. Roderick, and H. Hashemi, "An integrated ultra-wideband timed array receiver in 0.13 μm CMOS using a path-sharing true time delay architecture," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2834–2850, Dec. 2007.

- [30] S. Lo, I. Sever, S.-P. Ma, P. Jang, A. Zou, C. Arnott, K. Ghatak, A. Schwartz, L. Huynh, V. Phan, and T. Nguyen, "A dual-antenna phased-array UWB transceiver in 0.18- μm CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2776–2786, Dec. 2006.
- [31] H. Hashemi, and A. Hajimiri, "Concurrent multiband low-noise amplifiers—theory, design, and applications," *IEEE Trans. Microwave Theory & Tech.*, vol. 50, no. 1, pp. 288-301, Jan. 2002.
- [32] S. Jeon, Y.-J. Wang, H. Wang, F. Bohn, A. Natarajan, A. Babakhani, and A. Hajimiri, "A Scalable 6-to-18 GHz concurrent dual-band quad-beam phased-array receiver in CMOS," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2008, pp.186-187.
- [33] A. Ismail, and A. Abidi, "A 3-10GHz low-noise amplifier with wideband LC-ladder matching network," *IEEE J. of Solid-State Circuits.*, vol. 39, no. 12, pp. 2269-2277, Dec. 2004.
- [34] P. Ikalainen, "Low-noise distributed amplifier with active load," *IEEE Microwave and Guided Wave Letters.*, vol. 6, no. 1, pp. 7-9, Jan. 1996.
- [35] H. Hashemi, X. Guan, A. Komijani, and A. Hajimiri, "A 24-GHz SiGe phased-array receiver–LO phase-shifting approach," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 2, pp. 614–626, Feb. 2005.
- [36] D. M. Pozar, *Microwave Engineering*, 3rd ed., John Wiley & Sons, New York, 2005.
- [37] D. Parker and D. C. Zimmermann, "Phased arrays – Part I: Theory and architectures," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 3, pp. 678-687, Mar. 2002.

- [38] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed. New York: John Wiley & Sons, 1998.
- [39] R. J. Mailloux, *Phased Array Antenna Handbook*, 2nd ed. Norwood, MA: Artech House, 2005.
- [40] A. Hajimiri, H. Hashemi, A. Natarajan, X. Guan, and A. Komijani, "Integrated phased array system in silicon," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1637-1655, Sep. 2005.
- [41] J. Spradley, "A volumetric electrically scanned two-dimensional microwave antenna array," *IRE International Convention Record*, vol. 6, pp. 204-212, Mar. 1958.
- [42] D. Parker and D. C. Zimmermann, "Phased arrays – Part II: Implementations, applications, and future trends," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 3, pp. 688-698, Mar. 2002.
- [43] J. Paramesh, R. Bishop, K. Soumyanath, and D. Allstot, "A four-antenna receiver in 90-nm CMOS for beamforming and spatial diversity," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2515–2524, Dec. 2005.
- [44] L. C. Godara, *Smart Antennas*, CRC Press, Boca Raton, FL, 2004.
- [45] S.-S. Jeon, Y. Wang, Y. Qian, and T. Itoh, "A novel smart antenna system implementation for broad-band wireless communications," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 5, pp. 600–606, May. 2002.
- [46] T. Nishio, H.-P. Tsai, Y. Wang, and T. Itoh, "A high-speed adaptive antenna array with simultaneous multibeam-forming capability," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 12, pp. 2483–2494, Dec. 2003.

- [47] X. Li, S. Shekhar, and D. J. Allstot, "Gm-booster common-gate LNA and differential colpitts VCO/QVCO in 0.18- μ m CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2609–2619, Dec. 2005.
- [48] B. Razavi, RF Microelectronics, NJ: Prentice Hall, Upper Saddle River, 1998.
- [49] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July-October 1948.
- [50] J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.
- [51] Federal Communications Commission "Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems," First Report and Order, ET Docket 98-153, FCC 02-48, April 2002 (<http://www.fcc.gov>).
- [52] Code of Federal Regulations –Title 47-, Part 15: Radio Frequency Devices:
http://www.access.gpo.gov/nara/cfr/waisidx_07/47cfr15_07.html.
- [53] Y. Zhang, et al., "A 0.18 μ m CMOS dual-band UWB transceiver," *IEEE ISSCC Dig. Tech. Papers*, pp. 114–115, Feb. 2007.
- [54] F. S. Lee, et al., "A 2.5nJ/b 0.65V 3-to-5GHz Subbanded UWB Receiver in 90nm CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 116–117, Feb. 2007.
- [55] Y. Zheng, et al., "A 0.18 μ m CMOS 802.15.4a UWB transceiver for communication and localization," *IEEE ISSCC Dig. Tech. Papers*, pp. 118–119, Feb. 2008.
- [56] J. R. Bergervoet, et al., "A WiMedia-compliant UWB transceiver in 65nm CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 112-113, Feb. 2007.
- [57] J. R. Bergervoet, et al., "A WiMedia-compliant UWB transceiver in 65nm CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 112-113, Feb. 2007.

- [58] J. Yuan and C. Svensson, "High-speed CMOS circuit technique," *IEEE J. Solid-State Circuits*, vol. 24, no. 2, pp. 62-70, 1989
- [59] C. S. Vaucher, I. Ferencic, M. Locher, S. Sedvallson, U. Voegeli, and Z. Wang, "A family of low-power truly modular programmable dividers in standard 0.35- μm CMOS technology", *IEEE J. Solid-State Circuits*, vol. 35, no. 7, pp. 1039-45, 2000
- [60] M. Bohr, "A new era of scaling in an SoC world", *IEEE ISSCC Dig. Tech. Papers*, pp. 23-28, Feb. 2009.
- [61] Matlab online manual: <http://www.mathworks.com/>.